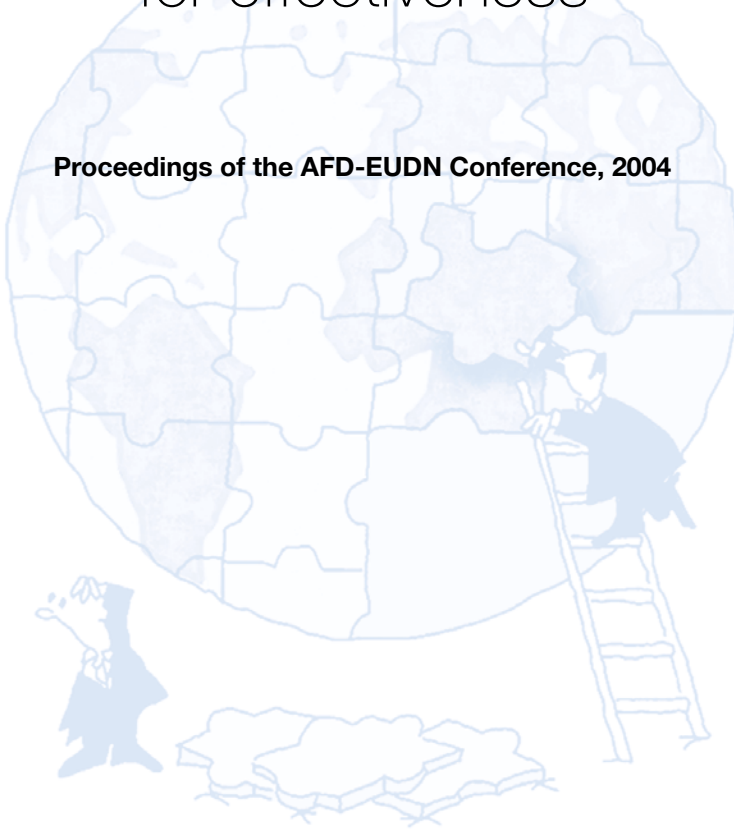
Notes and documents
Number 2
December 2005

Development Aid: Why and How?

Towards strategies for effectiveness

Proceedings of the AFD-EUDN Conference, 2004



Agence Française de Développement

Development Aid: Why and How? Towards strategies for effectiveness

Proceedings of the AFD-EUDN Conference, 2004

Research Department
Agence Française de Développement

Acknowledgements

We thank the members of the EUDN Network for their fruitful collaboration and the high quality of their contributions. We also express our gratitude to the French Ministry of Economy and Finance who agreed to host this event at the Pierre Mendès France Conference Centre. We thank all participants, whose presence largely contributed to the success of this second joint AFD-EUDN conference. We finally thank Jacky Amprou, Florence Avigni and Sandrine Karatchewsky-Volk for their precious help both in the organisation of the conference and the editing of this volume.

Nota:
La version française des communications de la deuxième conférence AFD/EUDN est publiée par la Revue d'Economie du Développement (No.2-3, Sept. 2005).

The French version of the second AFD/EUDN conference proceedings has been published in Revue d'Economie du Développement (No.2-3, Sept. 2005).

Disclaimer

The views expressed in this publication are those of the authors and do not necessarily reflect the official views of the Agence Française de Développement.

Table of contents

Introduction

Aide au développement : pourquoi et comment ?

Cet ouvrage rassemble les textes de la conférence co-organisée par l'AFD et le réseau de chercheurs européens EUDN (European Development Research Network) sur le thème « Aide au développement : pourquoi et comment – Quelles stratégies pour quelle efficacité ? ». Dans le cadre des débats actuels sur les Objectifs du millénaire pour le développement (OMD), cette conférence avait pour objet de contribuer à la réflexion sur le rôle de l'aide et les moyens d'accroître son efficacité pour atteindre les huit objectifs que la communauté internationale s'est fixés pour 2015.

Au cours des années 1990, l'aide au développement a traversé une crise de légitimité sans précédent, liée à la conjonction de plusieurs facteurs. Les mutations de l'environnement international, marqué alors par la fin de la guerre froide, ont transformé les fondements politiques de l'aide bilatérale, longtemps conçue comme un instrument de promotion des intérêts politiques et géostratégiques des pays donateurs. Parallèlement, les problèmes d'endettement croissants des pays receveurs et les crises financières successives ont ouvert un large débat sur la réforme de l'architecture financière internationale et sur le rôle assigné aux institutions multilatérales.

Enfin, les fondements économiques et les justifications de l'aide au développement ont été attaqués, avec la publication à la fin des années 1980 et au début des années 1990 d'études très critiques soulignant son absence d'efficacité macro-économique, ses effets potentiellement pervers pour les structures incitatives des pays en développement, les coûts sociaux et humains des ajustements structurels et l'échec des conditionnalités.

La remise en cause des fondements de l'aide, associée à la crise économique et aux contraintes budgétaires fortes pesant sur de nombreux pays donateurs, notamment les pays européens membres de la zone euro, ont entraîné à partir de 1992 une chute brutale des flux d'aide en direction des pays en développement. C'est dans cette conjoncture défavorable à l'aide internationale que la Banque mondiale a relancé le débat sur l'efficacité de l'aide et son allocation avec la publication de son rapport *Assessing Aid* (1998). Fondé sur les travaux de Burnside et Dollar, ce rapport soutient que l'efficacité de l'aide en matière de croissance dépend de la qualité des politiques économiques des pays en développement, ouvrant ainsi la voie au principe de sélectivité des pays receveurs sur la base de ce critère.

Parallèlement, les politiques d'aide au développement se sont recentrées sur l'objectif de lutte contre la pauvreté, recentrage concrétisé en 2000 par l'adoption aux Nations Unies des Objectifs du millénaire pour le développement – le premier de ces objectifs étant la réduction de moitié de la pauvreté dans le monde.

La deuxième conférence AFD/EUDN a permis de débattre de ces développements et de leur impact sur les politiques d'aide dans le cadre de l'engagement de la communauté internationale à atteindre les OMD en 2015. Les discussions se sont articulées autour de quatre questions : (i) Pourquoi aider les pays en développement ? L'aide est-elle l'outil le mieux adapté pour répondre aux problèmes des PED ? (ii) Comment allouer l'aide entre les pays receveurs ? Le modèle IDA fondé sur les performances doit-il être amendé et est-il applicable à l'aide bilatérale ? (iii) Quelles devraient être les priorités des politiques d'aide dans le secteur de la santé ? (iv) Pourquoi et comment évaluer l'impact des programmes d'aide au développement ?

Nous proposons dans cette introduction une synthèse des réponses apportées à ces questions lors de la conférence.

1. Pourquoi aider les pays en développement ?

Cette question a fait l'objet de la contribution de J. Gunning, qui l'a développée selon trois angles complémentaires. En effet, dans un contexte de restrictions budgétaires

dans les pays donateurs et de doutes quant à l'efficacité des politiques d'aide publique au développement, il paraît légitime de se demander : (i) quels sont les fondements actuels de l'aide au développement ? (ii) quels sont les arguments en faveur des politiques d'aide et (iii) l'aide est-elle l'outil le plus adapté pour répondre aux enjeux du développement des pays ?

1.1. Quels sont les fondements de l'aide aux pays en développement ? Une question dominée par le débat sur l'efficacité de l'aide

Si les discours des pays donateurs s'accordent tous sur la nécessité de promouvoir la croissance économique et de lutter contre la pauvreté, on observe dans les faits un décalage significatif entre les objectifs affichés par les bailleurs et les pratiques d'aide. Des travaux économétriques ont cherché à identifier les déterminants de l'allocation de l'aide par les différents bailleurs. Ils ont mis en évidence que l'allocation de l'aide dépend davantage de considérations politiques ou stratégiques que des besoins des pays receveurs ou de la qualité de leurs politiques. Cet écart entre les objectifs affichés de l'aide et son allocation réelle pourrait être un signe d'inefficacité. Cette idée est néanmoins nuancée par P. Guillaumont, qui remarque qu'il est en pratique difficile, voire impossible, de dissocier les motivations altruistes des intérêts des pays donateurs ou encore des intérêts des receveurs. Il existerait par ailleurs des divergences dans les objectifs poursuivis par les différents bailleurs. Par exemple, la pauvreté, la démocratie et l'ouverture seraient, selon les résultats des travaux économétriques, des déterminants plus importants pour l'aide nordique ou américaine que pour l'aide japonaise ou française.

Il est intéressant de noter que les débats récents, plutôt que de porter sur la question centrale des fondements de l'aide, se sont davantage concentrés sur celle, plus ciblée, de l'efficacité de l'aide. Pourtant, si ces deux questions sont fortement liées, celle de savoir pourquoi donner de l'aide ne pouvant être simplement ramenée à la démonstration de « pratiques » efficaces dans ce domaine.

Le débat sur l'efficacité de l'aide, largement influencé par les travaux de Burnside et Dollar, est aujourd'hui très présent chez les bailleurs et a fortement inspiré leurs modèles d'allocation ces dernières années. L'idée centrale est que l'aide a un impact positif plus important sur la croissance dans les pays menant de bonnes politiques

économiques. Ce résultat a donné naissance au concept de *sélectivité* consistant à cibler l'aide vers les pays *a priori* les mieux à même d'utiliser cette aide de manière efficace. Pourtant, comme l'ont souligné les différents intervenants, la sélectivité de l'aide suscite de nombreuses controverses. La première concerne les limites conceptuelles de cette approche. Gunning mentionne notamment l'absence de consensus autour de la notion de « bonnes politiques ». D'ailleurs, la littérature sur l'efficacité de l'aide a laissé de côté les considérations en termes d'économie politique, pour parler de « bonne gouvernance », concept plus large incluant les conditions de fonctionnement des institutions du pays receveur. La deuxième controverse concerne les limites des travaux académiques justifiant le recours à la sélectivité. P. Guillaumont précise qu'ils se cantonnent souvent à l'analyse de l'impact de l'aide sur la croissance, laissant de côté les autres dimensions du développement. De plus, la plupart se focalisent sur la qualité des institutions et des politiques des pays receveurs comme déterminant principal de l'efficacité de l'aide, alors que l'efficacité dépend aussi de facteurs tels que la vulnérabilité des pays. La troisième controverse concerne deux questions fondamentales peu débattues jusqu'alors et rappelées par J.-D. Naudet : i) aucune mesure empirique ne permet à ce jour de confirmer qu'une allocation sélective constitue en elle-même une incitation efficace à l'amélioration des politiques et des institutions des pays éligibles à l'aide ; ii) il peut y avoir un conflit entre la maximisation de l'efficacité de l'aide et des principes élémentaires de justice, notamment lorsque des facteurs exogènes (forte prévalence du VIH/SIDA par exemple) conditionnent également l'efficacité de l'aide.

Pourtant, malgré ces limites, les bailleurs restent largement favorables à l'extension d'une approche par la sélectivité.

1.2. Quels arguments en faveur de l'aide ?

La question de l'efficacité de l'aide ne suffit pas à justifier l'allocation d'une partie des budgets nationaux des pays développés à l'aide publique au développement. J. Gunning rappelle les principaux arguments en faveur de l'aide.

(i) L'aide permet de pallier les défaillances de marché, et notamment du marché du crédit. En effet, et ce malgré les rendements potentiels élevés des investissements, les PED rencontrent des difficultés à mobiliser des ressources

sur les marchés des capitaux. Gunning se penche en particulier sur la question du risque. Les chocs (commerciaux, climatiques, sanitaires, etc.) affectent les populations des pays du Sud de deux manières : *ex ante*, lorsque la perception du risque induit des changements des comportements d'investissements, et *ex post,* lorsque les chocs sont survenus. En l'absence de systèmes d'assurance, les ménages pauvres ont recours à des modes alternatifs de gestion du risque souvent très coûteux. Le marché n'offrant pas de systèmes d'assurance à ces populations, une action publique se justifierait donc dans ce domaine. Elle pourrait prendre la forme d'un partenariat public-privé dans lequel un bailleur assumerait une partie du risque de ré-assurance.

(ii) L'aide peut avoir des effets positifs sur la croissance en permettant aux pays receveurs d'alléger le poids de la fiscalité sur les activités productives, créant ainsi des incitations en faveur du secteur privé. Cet argument suscite néanmoins des controverses, de nombreux bailleurs souhaitant au contraire aider au renforcement de la fiscalité nationale, souvent embryonnaire.

(iii) L'aide peut inciter les gouvernements à mettre en place des politiques favorables à la croissance et à la réduction de la pauvreté. Pour aborder ces deux derniers points, J. Gunning fait appel à l'économie politique, notamment aux travaux d'Adam et O'Connell (1999). Ces derniers distinguent les régimes « prédateurs », cherchant à maximiser les transferts vers leur entourage, et les Etats « développeurs », plus démocratiques, où la taille du groupe « favorisé » a dépassé un certain seuil. Ils partent du postulat que l'aide constitue, avec la fiscalité, la principale source de revenu des pays receveurs. Ils notent également que les gouvernements ont tendance à vouloir effectuer des transferts vers les membres de leur entourage tout en essayant de fournir un minimum de services aux populations défavorisées afin de se maintenir au pouvoir. A partir de là, Gunning s'interroge sur la façon dont l'aide peut influencer ces différents types de régimes pour renforcer la croissance et lutter contre la pauvreté. Les conclusions de son analyse ne tranchent ni en faveur de la sélectivité, ni en faveur de la conditionnalité, qui a prouvé ses limites. L'échec de la conditionnalité s'explique notamment par l'absence de menace crédible d'interruption d'un programme : en effet, l'interruption de l'aide n'est pas dans l'intérêt des agences d'aide ; il s'agit par ailleurs d'un choix politiquement difficile pour un pays donateur. Pourtant, s'il

existait un véritable engagement des deux parties, et notamment une réelle menace de suspension de l'aide de la part des bailleurs, la conditionnalité constituerait dans un certain nombre de cas un instrument efficace.

(iv) Les externalités transfrontalières sont souvent identifiées comme des facteurs influant de façon positive ou négative la croissance. La santé est par exemple un bien public mondial ayant un impact fort sur le développement des pays du Sud. Depuis quelques années, le débat sur le rôle des pays donateurs, notamment par l'intermédiaire de l'aide, dans le financement de biens publics mondiaux est en cours.

Néanmoins, si les domaines d'intervention présentés par J. Gunning peuvent être consensuellement considérés comme prioritaires pour l'aide, J.-D. Naudet souligne qu'il existe d'autres priorités tout aussi valables : appui aux systèmes éducatifs ou aux systèmes de santé, protection de l'environnement, *empowerment* des populations les plus faibles, etc. Naudet rappelle que les pays receveurs ont en effet des priorités très différentes les uns des autres : il est donc fondamental d'être davantage à l'écoute de leurs préoccupations dans la définition des priorités de l'aide.

1.3. L'aide est-elle l'outil le plus adapté pour répondre aux problèmes des PED ? Quelles sont les autres actions envisageables ?

L'APD, entendue comme un soutien financier à un gouvernement, est-elle l'outil le mieux adapté pour atteindre les objectifs de développement visés ? Répondre à cette question implique d'établir un diagnostic clair de l'origine des problèmes auxquels sont confrontés les pays pauvres et de procéder à un calcul coût/avantage des différents outils à la disposition des pays donateurs. J. Gunning rappelle que, dans les faits, ce type de démarche n'est jamais mis en œuvre. Pourtant, il est possible d'imaginer que les donateurs mobilisent d'autres instruments que l'aide pour contribuer à relever les défis du développement. J. Pasquier a repris cette idée en plaidant pour une aide qui ne soit qu'un outil parmi d'autres politiques en faveur des PED.

J. Gunning évoque en particulier deux cas de figure où il existerait un intérêt majeur à mobiliser d'autres instruments que l'aide.

(i) Le premier concerne les petits pays, souvent africains (comme l'Ouganda ou le Mozambique), ayant récemment mis en place des réformes créant un environnement favorable au développement du secteur privé. Malgré ces efforts, ces pays n'ont toujours pas accès aux marchés des capitaux en raison d'une mauvaise diffusion de l'information à propos des réformes en cours. Pour Gunning, dans ce cas de figure, les pays donateurs ont un rôle d'informateurs à jouer, notamment au travers d'organismes multilatéraux comme le FMI, auprès des acteurs du secteur privé susceptibles d'investir dans ces pays.

(ii) Le second concerne les pays en situation de pré-conflit. Si l'aide a des effets importants sur la croissance des pays en situation de post-conflit, elle a aussi un rôle à jouer dans la prévention des conflits, qui trouvent souvent leur origine dans la pauvreté et contribuent à la renforcer. Mais en conjonction avec les politiques d'aide, d'autres actions devraient également être envisagées pour prévenir les conflits : réglementation du commerce de diamants pour rendre plus difficile le financement des rebellions, envoi d'opérations de maintien de la paix des Nations unies pour éviter l'explosion des violences, etc.

2. Comment allouer l'aide ? Critique du modèle IDA d'allocation fondée sur les résultats

Les travaux sur l'efficacité de l'aide ont eu pour prolongement logique un débat sur la prise en compte des facteurs d'efficacité dans les processus d'allocation de l'aide aux pays receveurs. Le papier de R. Kanbur s'inscrit dans cette lignée des travaux sur l'allocation optimale de l'aide. Il propose ici d'adapter un modèle d'allocation existant, en l'occurrence celui de l'IDA (International Development Association), en y introduisant des considérations telles que la prise en compte des performances des pays receveurs.

2.1. Les éléments du modèle IDA

Le modèle d'allocation de l'IDA permet d'allouer annuellement les fonds IDA disponibles pour les trois années à venir. Il comprend trois étapes. La première consiste à traduire les « performances » du pays sous la forme d'une note (*Country Performance*

Rating). La seconde étape intègre cette note dans une formule d'allocation qui donne une allocation triennale indicative par pays éligible à l'IDA, et la dernière ajuste cette allocation indicative en fonction de la stratégie d'aide pour déterminer l'allocation définitive.

Le Country Performance Rating

L'évaluation des performances des pays receveurs se fonde sur deux notations issues de deux évaluations internes à la Banque mondiale : le *Country Policy and Institutionnal Assessment* (CPIA) et le *Annual Report on Portfolio Performance* (ARPP). Le CPIA est un indicateur de qualité des politiques considérées par la Banque mondiale comme « favorisant la croissance économique et la réduction de la pauvreté ». Bien qu'étant un indicateur interne à la Banque mondiale, non publié, le CPIA a largement été repris par les travaux de la Banque sur l'allocation de l'aide en fonction des performances. Le texte revient sur ses modes de calcul et ses limites. L'ARPP, quant à lui, mesure les performances des opérations de l'IDA en cours dans le pays receveur, à partir du concept de « projets à risque »[1]. Le pourcentage de ces projets est converti en une note allant de 1 à 6, correspondant à l'échelle retenue pour le CPIA[2].

Ces deux indicateurs (CPIA et ARPP) sont ensuite combinés en un indicateur agrégé, selon une pondération 80:20 en faveur du CPIA, qui devient de fait l'élément dominant du *Country Performance Rating*. Pour renforcer encore l'accent mis sur la bonne gouvernance dans l'évaluation des performances du pays, une dernière étape dans le calcul du CPR consiste à multiplier l'indicateur agrégé par un « facteur de gouvernance ».

Prise en compte des besoins des pays receveurs

Sur la base du CPR, un niveau d'aide indicatif par habitant est calculé pour chacun des pays éligibles, tenant compte (i) du niveau de revenu du pays, permettant

1. Des discussions sont en cours pour substituer, dans le calcul du CPR, la notation *ex post* donnée par OED aux projets achevés dans l'année, à la notation ARPP, définie par les départements opérationnels sur les projets en cours.
2. L'IDA a reconnu les problèmes posés par cet indicateur, notamment ceux liés à l'augmentation de l'aide programme sous forme d'appui budgétaire, par rapport à l'aide sous forme de projet. Pour l'instant, le calcul de l'ARPP n'a pas été modifié pour prendre en compte cette évolution vers une approche programmatique de l'aide de l'IDA.

d'augmenter l'allocation en faveur des pays à bas revenu et (ii) d'un ensemble de corrections, permettant, par exemple, d'accorder davantage aux pays en situation de post-conflit[3] ou, au contraire, de fixer un maximum pour les pays également éligibles à la BIRD[4].

L'ajustement par le niveau de revenu est un élément important de ce modèle d'allocation, puisque c'est le seul qui prenne en compte les besoins du pays receveur (par rapport à ses performances), et donc qui introduise un biais en faveur des pays les plus pauvres. Il n'en reste pas moins que l'arbitrage entre performance et besoin se fait très largement en faveur de la première, puisque le CPR a un impact seize fois plus important sur le niveau d'allocation que le niveau de revenu.

Cette allocation indicative devient alors un point d'ancrage pour la « stratégie d'assistance pays » (CAS) de l'IDA, qui détermine le programme définitif des prêts au pays bénéficiaire. Deux scénarios (hypothèse haute et hypothèse basse) sont proposés et en fonction de l'évolution d'un certain nombre d'indicateurs précisés dans la CAS, l'un ou l'autre des scénarios sera finalement retenu. Cette démarche apporte une dimension dynamique au processus d'allocation des fonds IDA et procure des incitations au pays bénéficiaire pour améliorer ses performances pendant la durée de vie du fonds IDA. L'ampleur de l'écart entre l'allocation réelle par pays et l'allocation indicative du modèle est substantielle, puisqu'elle atteint en moyenne plus ou moins un tiers.

2.2. Le CPIA : un indicateur de performance ?

Le CPIA a pour objectif de quantifier les variables relatives à l'action gouvernementale et qui déterminent la croissance et la réduction de la pauvreté[5]. Le calcul de cet indicateur est réalisé depuis une vingtaine d'années et sert à noter les pays de l'IDA

3. Pour les pays en post-conflit où la capacité d'absorption le permet, l'allocation est à peu près multipliée par deux pendant les trois premières années suivant la fin du conflit.

4. Ces pays sont appelés « Blend countries ». Leur revenu par tête est inférieur à $875, mais leur capacité d'endettement est suffisante pour avoir accès aux crédits de la BIRD. Dès lors, leur accès aux ressources de l'IDA est plafonné, comme dans le cas de l'Inde (plafond à 2 milliards de DTS), du Pakistan (700 millions de DTS), de l'Indonésie (315 millions de DTS) et du Nigeria (750 millions de DTS).

5. L'indicateur est composé de 20 rubriques, également pondérées, et regroupées en quatre catégories principales : gestion macro-économique, politiques structurelles, politique en faveur de la réduction de la pauvreté et gestion du secteur public et des institutions. Chaque pays obtient une note individuelle sur une échelle allant de 1 à 6.

et de la BIRD. Les notes sont confidentielles et ne sont communiquées qu'au pays concerné[6].

Plusieurs critiques sont formulées à l'égard de cet indicateur. La première, rappelée par R. Kanbur, concerne l'absence de rubrique faisant référence aux résultats (*outcome*), tels que la croissance, le taux de mortalité infantile et/ou maternelle, l'éducation des filles ou toute autre dimension des ODM, puisque ces derniers peuvent être influencés par des facteurs exogènes, indépendants de la politique du gouvernement. En ce sens, le CPIA représente une évaluation des instruments (politiques et institutions) plutôt qu'une évaluation des performances des pays.

Par ailleurs, les évaluations sont censées être menées de manière cohérente, sans tenir compte du niveau de développement, ce qui va dans le sens d'une évaluation de l'effort (sujet à subjectivité) davantage que de l'*outcome*. Néanmoins, malgré l'importance donnée à l'effort, l'indice CPIA mesure toujours la qualité des politiques et des institutions en place, et non les intentions ou projets d'améliorations. De ce fait, le CPIA est une mesure statique de l'environnement politique et institutionnel. Et ce processus de notation ne permet pas de compenser les obstacles éventuels que certains pays peuvent rencontrer pour adopter et mettre en œuvre certaines politiques.

Sur ce thème, S. Michaïlof a lui aussi remis en cause le principe et le processus même de notation au sein de la Banque mondiale, qui manquent de souplesse et ne correspondraient pas à un véritable exercice d'évaluation des performances des pays emprunteurs.

2.3. Proposition d'amendement et critique du modèle IDA

Afin de pouvoir prendre en compte les performances du pays, R. Kanbur propose non pas de modifier la formule IDA, mais d'ajouter un critère au CPIA. Ce critère supplémentaire prendrait en compte l'évolution d'un indicateur d'*outcome*, lié à la réalisation d'un ou plusieurs ODM. Pour prendre en compte la performance, c'est-à-dire l'amélioration des politiques, le taux de variation de l'indicateur (calculé à partir d'une série temporelle) serait rapporté aux montants d'aide reçue et transformé en indice sur une échelle de 1 à 6 comme les autres composantes du CPIA.

6. Sous la pression du monde académique en particulier et parce qu'elle a utilisé le CPIA dans de nombreux travaux économétriques, la Banque mondiale a accepté de les rendre public à partir de 2006.

Cette modification, tout en conservant intacte la formule, et en particulier les pondérations, permet d'intégrer la notion de performance. Elle doit être perçue comme le premier temps d'un processus de révision du CPIA. En effet, N. Van de Walle a fait remarquer que l'introduction de cette catégorie supplémentaire ne changerait probablement pas considérablement l'allocation indicative issue du modèle IDA, mais permettrait d'intégrer une dimension absente jusqu'alors.

La portée limitée de la proposition fait également dire à N. van de Walle que, si l'objectif est de mettre en place un cadre incitatif, l'important pour les bailleurs est d'envoyer un signal cohérent aux pays receveurs ; pour ce faire, la coordination des bailleurs, plus que l'adoption d'un modèle d'allocation, apparaît comme la clé. M. McGillivray considère aussi que la sophistication du modèle IDA n'est pas la priorité et qu'étant donné le niveau de complexité du lien entre efficacité de l'aide et orientation des politiques économiques, l'essentiel est de mieux comprendre les déterminants de cette relation : la capacité d'absorption (effet de seuil probable), la non-linéarité, l'endogénéïté, etc.

Au-delà de ces aspects particuliers, la justification même d'un modèle d'allocation, avec toutes les rigidités qu'il implique, est discutée par S. Michaïlof et N. van de Walle. Même s'ils reconnaissent que l'adoption d'un tel modèle permet d'éviter à la Banque mondiale d'allouer l'aide en fonction de critères subjectifs et idiosyncrasiques, ils comparent le modèle à une « boîte noire » d'où sortirait une formule d'allocation magique. Les intervenants sont très sceptiques quant à l'opportunité de généraliser ce type de modèle à l'aide bilatérale. Ils s'interrogent également sur une série de questions connexes : la communauté des bailleurs doit-elle adopter un ensemble de critères d'allocation communs ? Des objectifs non liés directement au développement sont-ils acceptables ? Faut-il gérer les volumes d'aide en « bon père de famille » en suivant un modèle figé dans le temps, ou en « capital-risqueur » pariant sur les opportunités (S. Michaïlof) ?

3. Quel rôle pour les politiques d'aide en matière de santé ?

Le papier de C. Bell et C. Fink s'inscrit dans le débat sur l'allocation de l'aide au niveau non plus géographique, mais sectoriel : dans un contexte de ressources rares,

quelles devraient être les priorités des bailleurs ? Les auteurs considèrent que la lutte contre les maladies transmissibles est une priorité pour l'APD, pour trois raisons principales : (i) le poids de la maladie, chez les plus pauvres, est surtout le fait de maladies transmissibles ; (ii) lutter contre les maladies transmissibles est un « bien public » ; (iii) lutter contre les maladies transmissibles revêt également un caractère de bien public mondial qui justifie d'autant plus les flux d'aide du Nord vers le Sud. Les auteurs présentent les fondements théoriques justifiant ce choix en faveur de la lutte contre les maladies transmissibles et proposent des critères de décision qui pourraient guider les bailleurs dans la définition de leurs interventions.

3.1. Quels arguments en faveur de l'allocation de l'aide à la lutte contre les maladies transmissibles ?

La lutte contre les maladies transmissibles est un enjeu à la fois national, en particulier pour les PED les plus affectés, et international.

(i) Les maladies transmissibles touchent proportionnellement davantage les pays les plus pauvres et, au sein de ces pays, les populations les plus défavorisées. Selon les auteurs, le poids des maladies transmissibles, des maladies maternelles et infantiles ainsi que de la malnutrition serait proportionnellement nettement plus important dans les PED que dans les pays industrialisés : les maladies transmissibles représenteraient près de trois quarts du poids de la maladie en Afrique. Ces affirmations sont néanmoins nuancées par A. Soucat, qui invite à mieux différencier dans les calculs les maladies transmissibles *stricto sensu* des maladies « apparentées ».

(ii) Les maladies transmissibles ont un caractère « public » et leur contrôle a de fortes externalités positives. En particulier, les maladies transmissibles altèrent le niveau de capital humain actuel et, quand les malades sont de jeunes adultes, la formation du capital futur. L'allocation de l'aide à la lutte contre ce type de maladies peut donc contribuer à l'amélioration du bien-être et du capital humain de la population du pays récipiendaire. Là encore, A. Soucat apporte une légère nuance aux propos de C. Bell et C. Fink : le caractère « public » s'applique également à d'autres types de maladies.

(iii) Les pandémies sont enfin un « mal public mondial », dans le sens où elles ne peuvent être contenues au sein de frontières nationales et représentent une menace pour l'ensemble des pays. L'allocation de l'aide à la lutte contre ce mal public mondial trouverait alors sa justification dans l'intérêt propre des pays donateurs. Cet argument est un argument fort, même si, comme le souligne A. Soucat, il n'est encore une fois pas spécifique aux maladies transmissibles.

(iv) Par ailleurs, en raison de la faible capacité à payer des pays en développement, une grande partie de la R&D est actuellement concentrée au Nord et concerne les maladies touchant les populations des pays industrialisés. Soutenir la R&D sur des maladies transmissibles des pays en développement peut accroître l'intérêt des acteurs privés en leur faveur.

3.2. Comment lutter contre les maladies transmissibles ?

Afin d'identifier les actions les plus efficaces dans la lutte contre les maladies transmissibles, il est nécessaire d'évaluer l'ensemble des coûts engendrés par ces maladies. Pour C. Bell et C. Fink, il est important, dans l'identification des projets, de mesurer les coûts directs, indirects et potentiels—en cas de propagation—des diverses maladies transmissibles, aux niveaux individuel et national (effets sur le stock de capital humain et sur la formation du capital humain futur). Ils soulignent également l'importance des évaluations et de la diffusion de bonnes pratiques dans les interventions des bailleurs en faveur de la lutte contre ce type de maladies.

Quelles orientations de l'aide : prévention, traitement, R&D ?

Dans le débat sur l'allocation la plus efficace de l'aide en matière de lutte contre les maladies transmissibles (faut-il plutôt favoriser la prévention ou le traitement ? faut-il financer la R&D et dans quels domaines ?), les auteurs tranchent, sauf cas particuliers, en faveur de la prévention, d'autant plus que l'on se trouve dans un contexte de ressources rares. Cependant, la fongibilité, importante dans le secteur de la santé, compromet toute mesure précise de l'efficacité de l'aide. A l'opposé, dans la R&D du Sud, l'aide peut non seulement jouer un rôle décisif (mécanismes d'incitation, garanties d'achat), mais il est également possible de mesurer plus précisément son impact car,

comme le souligne J. Mathonnat, le risque de fongibilité est moindre dans ce domaine, les PED y consacrant très peu de ressources.

Une utilisation efficace de l'aide ?

Parallèlement à la question de l'allocation de l'aide en matière de santé se pose celle de l'efficacité de son utilisation. A ce titre, de nombreuses interventions ont souligné l'importance de la gouvernance et ont suggéré de consacrer une partie des financements au renforcement du cadre institutionnel dans ce secteur.

Vers la mise en place d'un programme global à long terme

La réflexion de C. Bell et C. Fink aboutit à une proposition opérationnelle : conformément aux conclusions qui ressortent de leur analyse, les auteurs suggèrent aux bailleurs de définir et financer un programme global de long terme visant à lutter contre les maladies transmissibles. Ce programme se déclinerait en plans nationaux, soutenus par l'assistance technique nécessaire et dotés d'un volet suivi-évaluation. Les résultats de ces plans, mesurés par un ensemble d'indicateurs, conditionneraient la totalité des prêts alloués au pays récipiendaire. La mise en place d'un tel programme exigerait de la part des bailleurs une forte concertation.

Ainsi, C. Bell et C. Fink présentent des arguments convaincants en faveur de l'allocation prioritaire de l'aide à la lutte contre les maladies transmissibles qui les amènent à proposer le financement par l'aide d'un programme global orienté vers cet objectif. Cependant, en concentrant l'APD sur ce seul type de maladies, d'autres enjeux sanitaires fondamentaux pourraient être négligés, comme le renforcement des systèmes de santé publics dans leurs dimensions institutionnelles, humaines et matérielles, renforcement pourtant indispensable à l'éradication des maladies transmissibles.

4. Pourquoi et comment évaluer l'impact global des programmes d'aide au développement ?

Malgré les sommes importantes injectées dans l'aide au développement chaque année, E. Duflo rappelle que nous savons encore peu de choses sur l'impact réel des

projets. Les interventions produisent-elles les bénéfices escomptés ? En connaît-on l'impact global sur la population ? Les programmes ou projets pourraient-il être mieux conçus pour atteindre les effets escomptés ? Les ressources sont-elles dépensées de façon efficiente ? Tels sont les questions auxquelles cherche à répondre l'évaluation d'impact d'un programme.

Aucun mécanisme de marché ne sanctionne la pertinence et l'efficacité des services apportés aux populations démunies par l'aide au développement. Dès lors, deux types de méthodes cherchent à évaluer l'efficacité des projets : i) l'évaluation opérationnelle (*Process Evaluation*), visant à vérifier le respect des procédures d'exécution et l'atteinte des objectifs prévus ; ii) l'évaluation d'impact, visant à mesurer l'impact d'un projet « toutes choses égales par ailleurs » et à identifier les bonnes pratiques qui découlent de l'évaluation.

E. Duflo (et, avec elle, F. Bourguignon) reconnaît que les évaluations opérationnelles sont indispensables et ont vocation à s'appliquer au maximum de projets. Elle précise qu'elles sont néanmoins insuffisantes : les évaluations d'impact doivent se développer. De plus, pour pouvoir modéliser les impacts d'un type de programme et en généraliser les enseignements, ces évaluations doivent se multiplier dans différents contextes.

4.1. Les méthodologies disponibles pour l'évaluation d'impact

La difficulté des évaluations d'impact réside dans l'impossibilité de comparer la situation d'un individu qui a bénéficié d'un programme et celle d'un individu qui n'en a pas bénéficié. La comparaison *avant–après* ne distingue pas les effets dus au projet de ceux dus à des facteurs exogènes au projet (autres programmes, réformes politiques, conjoncture économique, etc.). La comparaison *avec–sans* est également complexe, car la sélection intentionnelle des bénéficiaires biaise la comparaison entre populations bénéficiaires du programme et populations non bénéficiaires (biais de sélection). Plusieurs méthodes statistiques et économétriques permettent de reconstituer *ex post* des groupes de contrôle dont les caractéristiques s'approchent de celles du groupe de bénéficiaires. D'autres méthodes permettent de redresser *ex post* les biais de sélection, mais seule la *randomisation*[7] est vraiment fiable et rigoureuse pour mesurer

7. Ou « sélection par assignation aléatoire », méthode qui consiste à sélectionner de manière aléatoire, avant le lancement du programme, au sein du groupe de bénéficiaires potentiels, un « groupe témoin » qui aurait toutes les caractéristiques requises pour participer au programme mais n'y serait pas inclus.

l'impact réel d'un projet. En effet, elle seule permet d'identifier un groupe témoin en tout point similaire au groupe bénéficiaire, excepté sa non-participation au projet. Et si elle n'est pas infaillible (biais d'attrition, non prise en compte des externalités des programmes, etc.), ses défauts sont partagés par les autres types d'évaluation.

4.2. Champs d'application de la randomisation

La mise en œuvre de telles évaluations par randomisation est particulièrement adaptée à différents cas de figure : projet pilote avec groupes témoins, voué à se développer si l'évaluation est concluante ; phase de développement d'un programme existant ; exploitation des caractéristiques de certains programmes de type « loterie ». Ces situations évitent les questions éthiques et politiques qui peuvent être à l'origine du caractère non aléatoire de la sélection des bénéficiaires.

Bien sûr, tous les programmes ne peuvent être évalués par la technique de randomisation : l'aide budgétaire, ou les programmes de construction de grandes infrastructures, par exemple, ne s'y prêtent guère. En revanche, les projets qui servent des individus spécifiques ou des communautés et qui ont des objectifs bien définis dès leur lancement (cf. projet Progresa au Brésil) peuvent faire l'objet de telles évaluations.

4.3. Une évaluation d'impact rigoureuse offre une réelle valeur ajoutée

Les évaluations d'impact par randomisation sont encore relativement peu nombreuses, rarement rigoureuses et leur enseignements peu intégrés dans les pratiques d'aide.

Pourtant, elles offrent de nombreux avantages qui justifieraient leur montée en puissance : (i) elles permettent de mesurer l'impact réel du projet (au Kenya, par exemple, la distribution de livres n'a finalement d'impact que sur les meilleurs élèves) et donc (ii) de justifier la poursuite du financement d'un projet ou son extension ; (iii) de comparer, en termes de coût/avantages, les alternatives visant le même objectif ; (iv) d'améliorer les modalités de mise en œuvre d'un projet et (v) d'économiser des ressources rares. Même si ces évaluations prennent du temps, des enseignements partiels peuvent être obtenus en cours de réalisation du projet, et mieux vaut des enseignements fiables et précis tard que jamais. Le coût de ces évaluations

(principalement pour la collecte des données) est certainement inférieur aux gaspillages de ressources rares que leurs enseignements permettraient d'éviter. Elles permettent ensuite de (vi) procéder à une généralisation des enseignements, ce qui requiert beaucoup de prudence. A ce sujet, il faut rappeler que toute évaluation et, *a fortiori,* toute extension de programme reposent sur des fondements et un modèle théoriques que l'évaluation d'impact contribue à valider, notamment si l'évaluation s'effectue sur une grappe de projets similaires dans différents contextes. L'évaluation d'impact de projets pilotes, ajoute E. Sadoulet, peut servir à paramétrer de tels modèles structurels. Enfin, les évaluations d'impact peuvent permettre de contrer le scepticisme ambiant sur l'efficacité de l'aide, grâce à une connaissance étayée de ce qui marche ou pas en termes d'impact sur le développement.

L'évaluation d'impact rigoureuse serait en ce sens un bien public global (car participant de l'accroissement du stock commun de connaissances), dont la production en quantité et qualité suffisantes reposerait sur un partenariat entre organisations internationales (promotion et financement), ONG (capacité d'expérimentation), gouvernements locaux (internalisation des bénéfices de l'évaluation) et centres de recherche (formation, expertise technique).

L'initiative DIME (*development impact evaluation)* de la Banque mondiale, présentée par F. Bourguignon, met en œuvre cette méthode d'évaluation en l'appliquant à des grappes de projets pilotes concentrés sur quelques secteurs (infrastructures dans les bidonvilles, éducation primaire, allocations monétaires conditionnelles).

4.4. Les limites de la démarche

Cette démarche n'est pas dénuée de difficultés. Ainsi, H-R. Hemmer insiste, par exemple, sur les biais comportementaux que peut induire la connaissance d'une évaluation (modification des pratiques au sein des groupes de contrôle). L'identification de la « fonction objectif » est par ailleurs perçue par H-R. Hemmer et E. Sadoulet comme souvent délicate, en raison d'objectifs variables selon les acteurs.

Selon P. Jacquet, la difficulté de mise en œuvre de telles évaluations ne doit pas être sous-estimée, d'autant qu'elle se traduit par un coût élevé et que ce coût peut être excessif en regard des ressources allouées directement aux programmes eux-mêmes.

En réponse à la critique du coût élevé des évaluations par randomisation, une proposition d'E. Sadoulet consiste à davantage utiliser les données « administratives » collectées par l'entité projet. Une partie du financement pourrait être assumée par l'opérateur (ou le projet) si l'évaluation d'impact se transformait en un bien privé partiel générant des bénéfices à court terme pour le projet et l'équipe projet (effets d'apprentissage, d'amélioration de l'exécution et de capitalisation).

E. Sadoulet propose également deux pistes à explorer pour les évaluations d'impact : (i) élargir les dimensions de l'évaluation (règle d'éligibilité, calibrage des transferts, etc.) ; (ii) intégrer l'économie politique qui conditionne l'exécution du projet. Il est effectivement important de prendre en compte les institutions et le capital social locaux qui expliquent l'hétérogénéité des impacts d'une collectivité à l'autre.

5. Enseignements, interrogations et prolongements

Des différentes contributions à cette conférence, un certain nombre d'enseignements ressortent et des interrogations subsistent.

- L'aide ne constitue qu'un des outils à la disposition de la communauté internationale pour promouvoir le développement des pays moins avancés. Les politiques internationales de réglementation (du commerce en particulier), les politiques d'immigration, l'essor de sources alternatives de financement du développement sont autant de leviers complémentaires à manier de façon cohérente en faveur des pays en développement. L'avantage comparatif (ou efficacité relative) de chacun de ces leviers reste néanmoins à déterminer de manière plus précise en fonction des circonstances et des caractéristiques régionales.
- Les finalités de l'aide se sont multipliées : outil de gestion et de prévention des conflits, filet de sécurité sociale, outil de développement, outil de régulation de la mondialisation. Dans ce contexte, la question de la justification économique et politique de l'aide a souvent été assimilée à celle de son efficacité, alors même qu'un débat intense est toujours d'actualité sur les méthodes permettant d'évaluer cette efficacité. Au-delà de l'efficacité macroéconomique, l'évaluation de l'impact de l'aide au niveau microéconomique (impact des projets) apparaît

comme particulièrement critique dans les pays en développement où les ressources sont rares et où chaque dépense publique doit aspirer à maximiser son impact sur la réduction de la pauvreté.

- Les bailleurs de fonds ont de plus en plus recours à des modèles d'allocation de leur aide, fortement inspirés de celui de l'IDA et des travaux de Collier, Dollar et Burnside. Même si le recours à ce type de modèle permet d'éviter une allocation discrétionnaire et donc de renforcer la cohérence et la transparence des choix des donateurs, la question des critères reste posée. Un consensus semble se former au moins autour des deux critères que sont le niveau de pauvreté et les performances du pays receveur (malgré toutes les difficultés inhérentes à la mesure de telles performances). Néanmoins, cette liste n'est-elle pas trop restrictive ? Les critères ne devraient-ils pas être différents selon les bailleurs, en particulier entre multilatéraux et bilatéraux, dont les choix d'allocation peuvent être différents et complémentaires ? Enfin, dans un contexte d'annulation de dette (impliquant une allocation de l'aide en fonction d'engagements antérieurs), quelle place reste-t-il pour une allocation basée sur des critères de performance ?
- La lutte contre les maladies transmissibles doit être considérée comme une priorité des politiques d'aide en faveur du secteur de la santé, mais s'assurer que les conditions d'une utilisation efficace de l'aide dans ce domaine (renforcement des cadres institutionnels, humains et matériels de ce secteur) doit constituer également un souci majeur des bailleurs.

Au-delà des enseignements tirés des contributions, la conférence a aussi mis en lumière des axes de réflexion permettant de poursuivre les débats engagés. La perspective d'un doublement de l'aide en appui à la réalisation des OMD rend inévitable d'apporter des réponses à certaines questions, parmi lesquelles : (i) à qui allouer ce supplément d'aide ? et (ii) sous quelle forme ?

La réponse à la première question suppose de poursuivre les travaux sur l'efficacité de l'aide, en s'orientant vers une analyse plus fine de l'impact de l'aide sur les différents agrégats macroéconomiques. La notion d' « APD », telle que comptabilisée par le CAD regroupant des flux hétérogènes ayant des objectifs extrêmement variés, rend difficile d'en mesurer l'impact global sur la croissance ou la réduction de la pauvreté. Une approche plus appropriée semble être une désagrégation des flux d'aide et une analyse

de leur impact respectif sur les points d'action qui leur sont assignés. Les questions de la capacité d'absorption et des moyens de mesurer les besoins en aide apparaissent également comme un axe de recherche à approfondir. Une allocation optimale de l'aide dépend en partie du comportement des pays bénéficiaires et leur capacité à utiliser l'aide de manière productive. Or l'identification des déterminants de cette capacité d'absorption reste un champ d'investigation assez peu exploré jusqu'à présent.

Apporter des réponses à la seconde question, « sous quelle forme décaisser le supplément d'aide ? », nécessite de poursuivre la réflexion dans plusieurs domaines. Le premier relève du débat prêts ou dons, qui anime toujours la communauté des bailleurs de fonds. Au delà d'une analyse coûts-bénéfices de chacun des deux instruments, il s'agit de développer des complémentarités entre les deux et de proposer des montages financiers innovants adaptés aux spécificités des partenaires. Le deuxième concerne les conditions requises pour un accroissement de l'aide sous forme d'appui budgétaire, et en particulier celles liées à la fongibilité et au risque fiduciaire. Il existe désormais un consensus pour que ces aides budgétaires s'articulent autour des principes d'appropriation, d'alignement et d'harmonisation avec, comme contrepartie naturelle, une aide budgétaire basée sur une logique de résultats en termes de réduction de la pauvreté et un nécessaire renforcement des capacités nationales. Les conditions de réussite de l'aide budgétaire dépendent donc d'une meilleure connaissance et d'une amélioration des systèmes de gestion et de contrôle des dépenses publiques dans les pays receveurs. La question du renforcement des capacités dans ce domaine se pose tout particulièrement dans les pays dits « en situation de partenariat fragile ». Les donateurs sont toujours à la recherche d'approches appropriées pour aider les pays en situation de post-conflit ou plus généralement les pays disposant d'une structure étatique affaiblie. Enfin, un axe supplémentaire de réflexion concerne la volatilité de l'aide et son rôle en matière de compensation des chocs externes. Dans quelle mesure l'aide est-elle pro-cyclique et quelles en sont les implications en termes de politique économique ?

Introduction

Development Aid: Why and How?

This work brings together texts from the conference co-organised by AFD and EUDN (European Development Research Network) on the theme "Development Aid: Why and How? – Towards strategies for effectiveness." In the framework of current debates on the Millennium Development Goals (MDG), this conference has been organised to contribute to the reflection on the role of aid and ways to improve its effectiveness in order to reach the eight objectives set by the international community for 2015.

During the 1990s, development aid went through a crisis of legitimacy without precedent, due to various factors. Changes in the international environment, marked by the end of the Cold War, transformed the political foundations of bilateral aid, long considered as an instrument for promoting the political and geostrategic interests of donor countries. In the same way, problems of increasing debt in recipient countries and successive financial crises opened a broad debate over the reform of international financial architecture and the role given to multilateral institutions.

Finally, the economic foundations and the justification of development aid were attacked. During the late 1980s and early 1990s, some very critical studies emphasised the absence of macro-economic effectiveness, potentially perverse effects on incentive structures in developing countries, social and human costs of structural adjustment, and the failure of conditionalities.

The questioning of aid foundations—associated with the economic crisis and with strong budgetary constraints weighing on several donor countries, notably European countries in the Euro zone—has, since 1992, led to a brutal fall in aid flows to developing countries. It was under these unfavourable circumstances that the World Bank reopened the debate on aid and allocation effectiveness, with the release of its 1998 report entitled *Assessing Aid*. Based on work from Burnside and Dollar, the report supports

the idea that aid effectiveness in terms of growth depends on the quality of developing countries' economic policy, thus clearing the way for the principle of selectivity of recipient countries based on this criterion.

At the same time, development aid policies again focused on the fight against poverty. This focusing was emphasised in 2000 with the UN's adoption of the Millennium Development Goals—the first of these goals being the halving of world poverty.

The second AFD-EUDN conference facilitated a debate over these developments and on aid policy in the context of the international community's commitment to reach the MDGs by 2015. The discussions centred on four questions: (i) Why give aid to developing countries? Is aid the best tool for addressing the developing countries' difficulties? (ii) How should aid be allocated between recipient countries? Should the IDA performance-based model be amended, and is it applicable for bilateral aid? (iii) What priorities should aid to the health sector finance? (iv) Why and how should the impact of development aid programmes be evaluated?

In this introduction, we will propose a synthesis of the responses given to these questions during the conference.

1. Why give aid?

This question was addressed by Jan Gunning, who developed his response along three complementary angles. In the context of budgetary restrictions in donor countries, and along with doubts regarding the effectiveness of official development aid policy, it seems legitimate to ask: (i) what are the current foundations of development aid? (ii) what are the arguments in favour of aid policy? (iii) is aid the best tool for responding to countries' development issues?

1.1. What are the foundations of country development aid? A question dominated by the debate on aid effectiveness

While donor countries agree on the necessity of promoting economic growth and fighting poverty in their discourse, the facts show a significant discrepancy between

donors' objectives and aid practices. Some econometric studies have attempted to identify the determinants of aid allocation by donor. They showed how aid allocation depends more on political or strategic considerations than the needs of recipient countries or the quality of their policies. This gap between declared aid objectives and its real allocation could be a sign of ineffectiveness. This idea is nevertheless qualified by Patrick Guillaumont, who remarks that it is in practice very difficult, indeed impossible, to dissociate altruist motivations from the interests of donor countries or even recipient countries. Furthermore, different donors pursue different objectives. For example, according to econometric analyses, poverty, democracy and openness are more important determinants for Nordic and American aid than for Japanese or French aid.

It is interesting to note how recent debates have concentrated on the more specific question of aid effectiveness, rather than the central question of aid's foundations. Yet, while these two questions are strongly connected, the question "why give aid?" cannot be reduced to a simple demonstration of efficient "practices" in this area.

The debate on aid effectiveness, largely influenced by the works of Burnside and Dollar, is very present in the spirit of today's donors, and it has strongly inspired their allocation models in recent years. The central idea is that aid has a larger positive impact on growth in countries which have good economic policy. This result gave birth to the concept of *selectivity*, which consists in directing aid toward countries that are most likely to use this aid in an effective manner. However, as several participants have pointed out, aid selectivity brings up a number of controversies. The first concerns the conceptual limits of this approach. Gunning notably stresses the absence of consensus regarding the notion of "good policy". Furthermore, the literature on aid effectiveness has largely ignored political economy considerations, instead speaking of "good governance"—a broader concept that includes institutional characteristics of recipient countries. The second controversy concerns the limits of academic work justifying recourse to selectivity. Guillaumont points out that such often limits itself to an analysis of aid's impact on growth, leaving aside the other dimensions of development. And the majority of them focus on the quality of institutions and policy in recipient countries as the main determining factor of aid effectiveness—even though efficiency also depends on factors like a country's vulnerability. The third controversy concerns two fundamental questions, rarely debated until recently, and underlined by Jean-David

Naudet: (i) not a single empirical measurement available today allows us to confirm that a selective allocation constitutes, in itself, an effective incentive for countries eligible for aid to improve their policies and institutions, (ii) there can be a conflict between the maximisation of aid's effectiveness and elementary principles of justice, notably when exogenous factors (such as a high prevalence of HIV/AIDS) also condition the aid's effectiveness.

However, despite these limits, donors remain largely favourable to expanding the selectivity approach.

1.2. What arguments in favour of aid?

The question of aid effectiveness is not enough to justify allocating part of a developed country's national budget to official development aid. J. Gunning points out the main arguments in favour of aid.

(i) Aid allows for a softening of market failures, notably in the credit market. Developing countries, despite the potentially high returns on investments, have, in fact, trouble mobilising resources in the capital markets. Gunning leans toward the question of risk in particular. Crises (trade, climatic, health) affect populations of countries in the South in two ways: *ex ante*, when the perception of risk induces changes in investment behaviour, and *ex post*, when crises arise. In absence of insurance systems, poor households have no other recourse than alternative risk-management methods, which are often very costly. As the market does not offer insurance systems to these populations, public intervention is justified in this area. It could take the form of a public-private partnership in which a donor assumes part of the re-insurance risk.

(ii) Aid could have positive effects on growth by permitting recipient countries to lighten the fiscal load on productive activity, thus creating incentives that favour the private sector. This argument nevertheless creates controversy, as many donors prefer on the contrary to prop up national taxation, often in its infancy.

(iii) Aid can encourage governments to set up policies favourable to growth and poverty reduction. To address these two last points, J. Gunning calls on political

economy models, notably in the work of Adam and O'Connell (1999). They distinguish "predatory" regimes, which seek to maximise transfers to their entourage, from more democratic "developer" states, where the size of the "favoured" group has surpassed a certain threshold. They begin from the postulate that the aid makes up, along with taxation, the main income source for the receiving country. They also note that governments tend to direct funds toward their entourage, while providing a minimum of services to the disadvantaged population so as to stay in power. From this starting point, Gunning focuses on the ways in which aid can influence these different types of regimes to step up growth and fight poverty. The conclusions of his analysis support neither selectivity nor conditionality, which has found its limits. The failure of conditionality is explained by the absence of a credible threat with regard to interrupting a programme. In fact, stopping aid is not in the interest of the aid agencies. Furthermore, it is a difficult political choice to make for a donor country. Yet if there is a real commitment between the two parties—and notably a real threat of aid interruption by the donors—conditionality can be an effective instrument.

(iv) Cross-border externalities are often identified as factors that influence growth. Health, for example, is a global public good that has strong impact on the development of countries in the South. The debate on the role of donor countries in the financing of global public goods—notably through aid—has been under way for several years now.

Nevertheless, while a consensus could consider the areas of intervention presented by J. Gunning as a priority for aid, J.-D. Naudet stresses that other priorities are also important: support for education or health systems, environmental protection, empowerment of the weakest sectors of the population, etc. Naudet reminds that different recipient countries have, in fact, very different priorities. Paying better attention to their preoccupations while defining aid priorities is thus fundamental.

1.3. Is aid the best tool for responding to developing countries' problems? What other actions are conceivable?

Understood as financial support to a government, is ODA the best tool for reaching the selected development goals? Answering this question implies setting up a clear

diagnosis of the origins of the problems with which the poor countries are confronted, and then calculating the cost/benefit of different tools at the donor countries' disposal. J. Gunning reminds us that, in fact, this type of approach is never implemented. But it is possible to imagine donors mobilising instruments other than aid in order to meet the challenges of development. J. Pasquier took up this idea by supporting an aid that is one policy tool among others that favour developing countries.

J. Gunning evokes in particular two scenarios where there is a major interest in mobilising instruments other than aid.

(i) The first concerns the small, often African, countries (like Uganda or Mozambique) which have recently implemented reforms creating favourable environments for the development of the private sector. Despite their efforts, these countries still lack access to capital markets because of a poor dissemination of information regarding the reforms under way. For Gunning, in this scenario, donor countries have an informative role to play, notably through multilateral organisations like the IMF, and with the private sector actors who are likely to invest in these countries.

(ii) The second concerns the countries in pre-conflict situations. While aid has important effects on growth in post-conflict countries, it also has a role to play in the prevention of conflicts, which are often caused and perpetuated by poverty. Yet, in conjunction with aid policy, other actions should also be envisioned to better prevent conflicts: regulating the diamond trade, for example, to make financing rebellions more difficult, sending United Nations peace-keeping missions to prevent an explosion of violence, etc.

2. How allocate aid? A critique of the IDA results-based allocation model

The work on aid effectiveness has led to a debate over taking effectiveness factors into account in the aid allocation process. Ravi Kanbur's paper is in line with these works on optimal aid allocation. He proposes adapting an existing allocation model, specifically that of the IDA (International Development Association), by introducing considerations such as the performance of recipient countries.

2.1. The elements of the IDA model

The IDA aid allocation model allows for an annual allocation of IDA funds available for the three years ahead. It consists of three stages. The first stage sees country "performance" translated into a grade (*Country Performance Rating*). In the second step, this grade is integrated into an allocation formula, which indicates a triennial allocation for countries eligible for the IDA. The final step adjusts the indicated allocation according to the aid strategy, so as to determine the definitive allocation.

The Country Performance Rating

Evaluation of recipient countries' performance is based on two grading systems, which stem from two internal evaluations at the World Bank: the *Country Policy and Institutional Assessment* (CPIA) and the *Annual Report on Portfolio Performance* (ARPP). The CPIA is an indicator of the quality of policy considered by the bank to "favour economic growth and poverty reduction". Despite being an undisclosed internal indicator at the World Bank, the CPIA was largely used in the Bank's work on aid allocation according to performance. The text goes back over its calculation methods and their limits. As for the AARP, this measures the performance of IDA operations currently running in the recipient country, starting from the concept of "risk projects".[1] The percentage of these projects is converted into a grade ranging from 1 to 6, which corresponds to the scale reserved for the CPIA.[2]

These two indicators, (CPIA and ARPP) are then combined into an aggregate indicator, with weighting of 80:20 in favour of the CPIA, which becomes the dominant element in the *Country Performance Rating*. To further reinforce the stress put on good governance in the evaluation of country performance, a last stage in the calculation of the CPR consists of multiplying the aggregate indicator by a "governance factor".

1. Regarding CPR calculations, discussions are under way to replace the ARPP grade, defined by operational departments and based on projects in progress, by the *ex post* grade—given by OED to projects closed during the year.

2. IDA has aknowledged the problems raised by this indicator, in particular those linked to increases in programme aid via budgetary support, in comparison with project aid. At the moment, the calculation of ARPP has not been modified to take into account this evolution toward a programmatic approach of the aid from IDA.

Taking into account the needs of recipient countries

Based on the CPR, the level of aid per inhabitant is calculated for each of the eligible countries, taking into consideration (i) the income level of the countries, allowing for increases in allocation for low income countries; (ii) a set of corrections—which allows, for example, to be given higher grants to countries in post-conflict situations[3], or maximum limits to be set for those countries also eligible for the IBRD.[4]

The adjustment by income level is an important element of this allocation model because it is the only one that takes into account the needs of the recipient country (in relation to its performance) and thus introduces a way to favour the poorest countries. Even so, the arbitrage between performance and need remains, and this is largely in favour of the former, because the CPR has an impact 16 times greater on the allocation level than the income level.

This indicative allocation thus becomes a cornerstone for the IDA's "country assistance strategy" (CAS), which determines the definitive programme of loans to beneficiary countries. Two scenarios (high- and low- hypotheses) are proposed, and according to the evolution of a certain number of indicators specified by the CAS, one of the scenarios will be ultimately chosen. This step brings a dynamic dimension to the process of IDA aid allocation while giving incentives to beneficiary countries to better their performance during the lifetime of the IDA funds. The amplitude of this gap between real allocation by country and the model's indicative allocation is substantial, because it reaches, on average, more or less a third.

2.2. The CPIA: an indicator of performance?

The CPIA's objective is to quantify the variables relative to government action, which determine growth and poverty reduction.[5] The calculation of this indicator is done

3. For post-conflict countries where the absorption capacity allows for it allocation is multiplied approximately by 2 during the first 3 years following the end of the conflict.

4. These countries are called "Blend Countries". Their per-capita income is lower than $875, but their debt capacity is sufficient to allow for access to IBRD credits. From that point on, a ceiling is set regarding their access to IDA resources, for example as with India, (a ceiling of 2 billion DTS), Pakistan (700 million DTS), Indonesia (315 million DTS), and Nigeria (750 million DTS).

5. The indicator is composed of 20 columns, weighted equally and grouped into four main categories: macroeconomic management, structural policy, pro-poverty reduction policy, and management of the public sector and institutions. Each country receives an individual grade on a scale of 1 to 6.

over an approximately 20-year period, and it serves to grade IDA and BIRD countries. The grades are confidential and transmitted only to the countries concerned.[6]

Several critiques are addressed to this indicator. The first, evoked by R. Kanbur, concerns the absence of a column that refers to results (outcome) such as growth, infant or mother mortality rates, female education, or any other dimensions of the MDGs, because these can be influenced by exogenous factors independent of the government. In this sense, the CPIA represents an evaluation of instruments (political and institutional) rather than an evaluation of country performance.

In other respects, the evaluations are supposed to be conducted in a coherent manner, without taking into account the development level, —something that leans more toward an evaluation of effort (subjectivity subject) than outcome. Nevertheless, despite the importance given to effort, the CPIA index always measures the quality of policies and institutions in place, not intentions or improvement projects. Because of this, the CPIA is a static measure of the political and institutional environment. Moreover, this grading process does not allow for compensating for the obstacles some countries may meet when adopting and implementing certain policies.

On this theme, S. Michailof has also questioned the principle and even the World Bank process of grading, which is lacking flexibility and does not suit a real evaluation of recipient countries' performance.

2.3. Critique of the IDA Model and a proposal for changing it

In order to take into account country performance, R. Kanbur proposes not to modify the IDA formula, but to add a criterion to the CPIA. This new criterion would take into account the evolution of an outcome indicator, linked to the realisation of one or more MDGs. To take performance into account, that is to say, the improvement of policy, the rate of variation of the indicator (calculated from a temporal series) would be divided by the amounts of aid received and transformed into an index on a scale of 1 to 6, like the other components of the CPIA.

6. Notably under pressure from the academic community and also because it has used the CPIA in many econometric studies, the World Bank has agreed to make the grades publicly available in 2006.

This modification permits to be integrated, the notion of performance while keeping the formula and the weighting unchanged. It should be seen as a first revision of the CPIA. In effect, N. van de Walle has emphasised that the introduction of this additional category would most likely not change the IDA model's indicative allocation, but would allow for the integration of a dimension that had until now been missing.

The limited scope of the proposition also makes it clear to N. van de Walle that while the objective is to put in place an incentive framework, sending a coherent signal to recipient countries is what is important to donors. To do this, donor co-ordination, more than the adoption of an allocation model, appears key. M. McGillivray also considers that the sophistication of the IDA model is not a priority, and that given the complexity of the link between aid effectiveness and orientation of economic policy, what is essential is better understanding the determining factors of this relation: absorption capacity (probable threshold effect), non-linearity, endogeneity, etc.

Beyond these particular aspects, S. Michaïlof and N. van de Walle discuss the very justification of an allocation model—with all the rigidities that this implies. While they recognise that the adoption of such a model allows the World Bank to avoid allocating aid according to subjective and idiosyncratic criteria, they compare the model to a "black box" from which a magic allocation is spawned. Participants are very sceptical as to the possibilities of using this type of model on bilateral aid. They are also asking themselves a series of related questions: should the donor community adopt a common set of allocation criteria? Are objectives that are not directly linked to development acceptable? Should one manage aid volume in a conservative manner, following old (maybe outdated) models, or as a "venture capitalist", betting on opportunities (S. Michaïlof)?

3. What role for aid policy with regard to the health sector?

C. Bell and C. Fink's paper falls within the scope of the debate over an aid allocation that is no longer geographical, but sectorial. In an environment where resources are rare, what should donors consider as priority? The authors think that the fight against communicable diseases is a priority for ODA for three main reasons: (i) the bulk of disease, among the poorest, is by far communicable disease; (ii) fighting against

communicable disease is a "public good"; (iii) fighting against communicable disease also assumes the characteristics of a "global" public good, which justifies even more the flow of aid from the North to the South. The authors present the theoretical basis that justifies this choice, which favours the struggle against communicable diseases, and propose decision-making criteria that can guide donors in defining their interventions.

3.1. What arguments for allocating aid to the fight against communicable diseases?

The fight against communicable diseases is an issue that is both national—in particular for the most affected developing countries—and international.

(i) Communicable diseases affect, proportionally, more often the poorest countries —and within these countries, the most disadvantaged populations. The authors say that the weight of communicable disease, maternal and infant disease as well as malnutrition will be clearly greater in the developing countries than in industrialised countries. Communicable diseases represent almost three-quarters of disease in Africa. A. Soucat nevertheless makes a fine distinction of these assertions, proposing a differentiation between *stricto sensu* communicable diseases and diseases that are "related".

(ii) Communicable diseases have a "public" trait and controlling them has strong positive externalities. In particular, communicable diseases alter the level of current human capital, and, when the sick are young adults, the training of future capital. In the fight against this type of disease, aid allocation can thus contribute to the improvement of well-being and human capital among the populations of recipient countries. Here again, A. Soucat brings a slight nuance to the remarks of C. Bell and C, Fink: the "public" trait also applies to other types of diseases.

(iii) In short, pandemic diseases are a "global public ill" in the sense that they cannot be contained within national boarders and they represent a threat for all countries. Aid allocation in the fight against this global public ill is justified in the donor countries' own interest. This argument is a strong one, even if, as A. Soucat stresses again, it is not yet specific to communicable disease.

(iv) Furthermore, because of the developing countries' weak capacity to pay, a large part of R&D is currently in the North and concerns diseases that affect the industrialised countries. Supporting the R&D for communicable diseases in developing countries can raise the private sector's interest in them.

3.2. How should communicable diseases be fought?

To identify the most effective actions in fighting communicable diseases, evaluating all the costs they generate is a must. C. Bell and C. Fink think that the direct, indirect and potential (in case the disease spreads) costs of several communicable diseases should be measured. And this should be done on both individual and national levels (the effects on the human capital supply and the training of future human capital). They also emphasise the importance of evaluating and communicating good donor practices in the fight against these types of diseases.

What directions for aid: prevention, treatment, R&D?

Should aid focus on prevention or treatment? Should R&D be financed, and if so, in what areas? These are the central questions in the debate on effective aid allocation for fighting communicable diseases. Except in certain cases, the authors favour prevention—and this even more in situations where resources are rare. However, fungibility—considerable in the health sector – compromises any precise measurement of aid effectiveness. On the other side of things, not only can R&D in the South play a decisive role (incentive mechanisms, purchase guarantees), its impact can also be measured more precisely, because, as J. Mathonnat points out, the risk of fungibility is less in this area as the developing countries dedicate few resources to it.

An effective use of aid?

Along with the question of aid allocation to the health sector, comes that of how to use it most effectively. On account of this, many interventions have stressed the importance of governance, and have suggested dedicating part of financing to the institutional framework of the sector.

Toward a long term global programme

C. Bell and C. Fink's reflection led to an operational proposal: in accordance with the conclusions of their analysis, the authors recommend donors to define and finance a long-term global program to fight communicable diseases. This programme would come in the form of national plans supported by necessary technical assistance and equipped with a monitoring/evaluation aspect. The results of these plans, measured by a set of indicators, would condition the totality of loans allocated to recipient countries. The setting up of such a program demands a strong donor dialogue.

C. Bell and C. Fink thus present some convincing arguments in support of prioritising aid allocation for the fight against communicable diseases. And this leads them to propose another idea: that aid be used to finance a global programme aimed at this goal. But by concentrating ODA on this type of disease only, other fundamental health issues could be neglected, like the capacity building in the public health sector, on the institutional, human and material levels. These reinforcements are indispensable to the eradication of communicable diseases.

4. Why and how should the global impact of development aid programmes be evaluated?

Despite the considerable amounts spent on development aid annually, E. Duflo points out that we still know little as to the real impact of the projects. Do interventions produce the expected results? And do we know what global impact they have on populations? Could the programmes or the projects be better conceived to produce the desired effects? Are resources spent in an effective manner? Such are the questions that the impact evaluation programme tries to address.

No market mechanisms can sanction the relevance and effectiveness of services brought by development aid to deprived populations. Two types of methods seek to evaluate the effectiveness of the projects: i) operational evaluation (*Process Evaluation*), which aims at verifying the respect of execution procedures and the fulfilment of expectations; ii) impact evaluation, which aims at measuring a project's impact with "all things being equal", and identifying the good practices that come from the evaluation.

E. Duflo (and along with her, F. Bourguignon) recognises that operational evaluations are indispensable and should be applied to a maximum of projects. She specifies that they are, nevertheless, insufficient: impact evaluation must be developed. What is more, in order to model the impacts of a type of programme, and generalise the lessons, these evaluations must be carried out more often in different contexts.

4.1. Methods available for impact evaluation

The difficulty in impact evaluations lies in the impossibility of comparing an individual who has benefited from a program with an individual who has not. The *before-after* comparison does not distinguish effects due to the project from those due to exogenous factors (other programmes, political reform, the economic situation, etc.) The comparison *with-without* is also complex, because the intentional selection of beneficiaries biases the comparison between populations that benefit from the programme and populations that do not (selection bias). Several statistical and econometric methods make it possible to reconstitute, *ex post*, the control groups whose characteristics come near those of the beneficiary groups. Others allow for an *ex post* rectification of the selection bias, but only *randomisation*[7] is really reliable and rigorous enough for measuring a project's real impact. In effect, it alone makes it possible to identify a witness group similar in every respect to the beneficiary group—aside from its non-participation in the project. It is not infallible (attrition bias, non-accounting of programme externalities, etc.), but its faults are also found in other types of evaluation.

4.2. Fields of application for randomisation

The implication of such evaluations by randomisation is particularly adapted to different scenarios: pilot projects with witness groups destined to develop if the evaluation is conclusive, development phases of existing programmes, and exploitation of the characteristics of certain "lottery" type programmes. These situations avoid the ethical and political questions that can cause the beneficiary selection process to take up a non-random character.

7. Where "selection by random allocation" is a method that consists of randomly picking a "control group" before the programme launch. This group would lie within the group of potential beneficiaries, and have all the required characteristics for participation in the programme, though it will not be included.

Of course, all programmes cannot be evaluated by the randomisation technique: budgetary aid or large infrastructure construction projects, for example, hardly lend themselves to it. On the other hand, projects that serve specific individuals or communities that have clearly defined objectives from the beginning (such as the Progresa project in Brazil) can be subjected to such evaluations.

4.3. Rigorous impact evaluation offers real added value

Impact evaluations by randomisation are still relatively few and rarely rigorous, and the lessons that they bring are still rarely integrated in aid practices.

However, they offer numerous advantages that justify their rise in importance: (i) they allow for a measurement of the project's real impact (in Kenya, for example, book distribution, in the end, only has an impact on the best students) and thus (ii) they justify the pursuit of financing a project or expanding it; (iii) they allow comparisons, in cost/benefit terms, of alternatives aiming for the same objective; (iv) they allow project implementation methods to be improved; (v) they allow to be saved rare resources. Even if these evaluations take time, some lessons can be learned during project realisation, and reliable and precise lessons come better late than never. The cost of evaluations (mainly for data collection) is certainly less than the waste of rare resources that lessons learned can help to avoid. They then permit (vi) lessons to be generalised—something that requires much prudence. As for this, one must remember that every evaluation, and *a fortiori* every programme expansion, rests on a foundation and a theoretical model that the impact evaluation helps validate—in particular, if the evaluation takes place on a group of similar projects in different contexts. The evaluation of pilot programme projects, E. Sadoulet adds, can serve to define such structural models. Finally, impact evaluations can help counter the prevailing scepticism on aid effectiveness, thanks to the fact that lessons can be confirmed, once it is known what works in terms of impact on development.

Rigorous impact evaluation would be in this sense a global public good (because it helps increase the common stock of knowledge). Its quantitative and qualitative production would reside in a partnership between international organisations (promotion and financing), NGOs (experimentation capacity), local governments (evaluation benefit internalisation), and research centres (education, technical expertise).

The World Bank's DIME initiative (*development impact evaluation*) presented by F. Bourguignon makes use of this evaluation method by applying it to clusters of pilot projects concentrated in a few sectors (infrastructures in shanty towns, primary education, conditional monetary allocations).

4.4. Limits to the approach

This approach does not lack its difficulties. H.-R. Hemmer for example, insists on the behavioural bias that can be brought on from the knowledge that an evaluation is occurring (changes in practice within the control groups). Moreover, the identification of the "objective function" is perceived by H.-R. Hemmer and E. Sadoulet as being often delicate, because objectives can change according to actors.

For P. Jacquet, the difficulty of implementing such evaluations should not be underestimated, all the more so because they translate into a higher cost and this cost can be excessive regarding resources directly allocated to the programmes themselves.

In response to the high cost of evaluation by randomisation, E. Sadoulet proposes improving the use of "administrative" data collected by the project entity. Part of the financing could be taken up by the operator (or the project), if the impact evaluation transforms into a partially private good that generates short-term benefits for the project and the project team (effects due to training, or improvement of execution and capitalisation).

E. Sadoulet also proposes two paths for exploration with regard to impact evaluation: (i) expanding the evaluation's dimension (eligibility rules, transfer calibration, etc.); (ii) integrating political economy models which condition the project's execution. In fact, it is important to take into account institutions and local social capital that explain impacts' heterogeneousness from one group to another.

5. Lessons, questions and repercussions

The different contributions to this conference have brought out a certain number of lessons, though questions remain.

- Aid is but one tool available to the international community for promoting development in the least advanced countries. International regulatory policy (trade in particular), immigration policy and rapid development of alternative financing sources for development are also complementary policy tools for helping developing countries in a coherent manner. The comparative advantage (or relative effectiveness) of each of these tools remains nevertheless to be determined in a more precise manner according to the regional circumstances and characteristics.
- The purposes of aid are diversified: it is a tool for the management and prevention of conflicts, a social safety net, a development tool and a tool for regulating globalisation. In this context, the question of economic and political justification of aid has often been assimilated to that of its effectiveness, while methods that allow for an evaluation of its effectiveness are constantly under intense debate. Beyond macroeconomic effectiveness, aid impact evaluation on the microeconomic level (project impact) appears particularly critical in developing countries where resources are rare and where public expenditure should aim to maximise its impact on poverty reduction.
- Donors have more and more recourse to allocation models for their aid, which are greatly inspired by those of the IDA and the works of Collier, Dollar and Burnside. Even if the use of this type of model avoids a discretionary allocation, thus reinforcing coherence and transparency in donor choice, the question of criteria remains open. There appears to be a consensus around two criteria at least: poverty level and recipient country performance (despite all the inherent difficulties of measuring such performances). Nevertheless, is this list not too restrictive? Should the criteria not be different according to the donors, in particular between multilateral and bilateral donors, whose allocations can be different and complementary? Finally, in a context of debt cancellation (implying aid allocation according to past engagements), what place remains for allocation based on performance criteria?
- The fight against communicable diseases should be considered a priority for aid policy in the health sector. But donors should also focus on the conditions for an effective use of aid in this area (reinforcing institutional, human and material frameworks in this sector).

Beyond the lessons learned from the contributions, the conference also brought light to the issues to work on in order to carry on the running debates. The prospect

that aid may be doubled in support of the MDGs makes certain questions inevitable, such as: (i) to which countries should this additional aid be allocated? and (ii) in what form should it be allocated?

The response to the first question assumes that the work on aid effectiveness will be continued, moving towards a finer analysis of aid impact on different macroeconomic indicators. The notion of "ODA"—as defined by the DAC—groups together heterogeneous flows that have extremely varied objectives, making a measurement of their global impact on growth or poverty reduction difficult. A more appropriate approach would appear to be one where aid flows are broken up and an analysis of their impact stays respective to the points of action they are assigned. Questions on absorption capacity and the means for measuring aid needs also appear to make up a line of research to be further developed. Optimal aid allocation depends in part on the behaviour of beneficiary countries and their capacity to use aid in a productive manner. Yet identifying the determinants of this absorption capacity remains a field of investigation that has been rarely explored to date.

Responding to the second question—"in what form should additional aid be distributed?"—requires doing more research in different areas. The first falls within the debate over loans or grants, something that always stirs up the donor community. Beyond a cost/benefit analysis of each of the instruments, it is a question of developing the complementary aspects between the two, and proposing innovative financial instruments adapted to partners' specificities. The second area concerns the conditions required for an increase in aid in the form of budgetary support, and in particular those linked to fungibility and fiduciary risk. From then on, there exists a consensus for these budgetary aids to structure themselves around the principles of appropriation, alignment, and harmonisation. The obvious counterpart of these principles is a budgetary aid where emphasis is based on results in terms of poverty reduction and a necessary reinforcement of national capacities. The conditions of success for budgetary aid thus depend on a better knowledge of, and an improvement of, management systems and public spending control in recipient countries. The question of capacity building in this area comes up particularly in countries with "fragile partnership situations". Donors are always looking for approaches that are appropriate for helping countries in post-conflict situations or, more generally, countries with a weak institutional structure. Finally, another line of reflection concerns aid volatility and its role in compensating for external crises. To what extent is aid pro-cyclic, and what are the implications of this in terms of political economy?

Why Give Aid?

by
Jan Willem Gunning
Free University, Amsterdam, AIID, CERDI, CSAE and EUDN

1. Introduction

Once upon a time (at least a quarter of a century ago), life in the world of development assistance seemed uncomplicated: there was very little agonising over the rationale for aid. Aid was, of course (just as now), partly perceived as an instrument of foreign policy and it was supposed to further the donor's (narrowly defined) self-interest. Yet, empathy (or altruism) also played a role: the welfare of developing countries was present in the donor's objective function. These arguments alone seemed sufficient justification for aid, for it was widely accepted that development required a combination of external investment finance and technical assistance. Obviously, donors could play a key role in providing these.

Though dissent always existed, notably from Peter Bauer, for decades it remained peripheral. Questioning why rich countries should give aid would have been considered odd, both amongst the experts and in civil society: the answer was considered obvious. There was always considerable debate regarding *how much* aid was needed and the suitable modalities for delivering it, but there was relatively little soul-searching on the fundamental justification for aid.

The Netherlands provides a case in point. For several decades (ending in the early 1990s), there was a very active domestic political debate on development. Yet, the key question in that debate always seemed to be what percentage of GDP should be allocated to development assistance. To raise the issue of how that money was to be used would have been interpreted as questioning the moral imperative of giving aid; an issue that was hardly a topic of debate in polite company.

Since then, this self-confidence of the aid community has evaporated: the rationale for aid has ceased to be self-evident. Aid fatigue set in when it appeared that, despite large amounts of aid, many poor countries were making very little progress. Critics argued that aid was counter-productive in that it supported governments that were hostile to economic growth and poverty reduction.

One way to interpret the question "why give aid?" is as a question about aims. In this light, the issue is what objective the donor tries to achieve by giving aid. The question is factual and the answer is relatively straightforward. The US, which by 1960 accounted for well over half of total development assistance, was largely motivated by political self-interest. Aid was expected to promote development and political stability and this would be in the strategic interest of the US. Humanitarian motives were more important for the smaller donors, notably the Scandinavian countries and the Netherlands.[1]

One can try to distil objectives from policy pronouncements but this has obvious limitations. A more convincing approach uses regression analyses to investigate what determines (in a statistical sense) aid allocations of various donors. Alesina and Dollar (2000) used this approach. They found that there is a wide gap between donor rhetoric and practice. Whether a country gets aid depends not so much on its needs and policies (if the country is poor, or if it has adopted "good" policies) as on political and strategic considerations, such as the country's colonial past and its voting behaviour in the United Nations. The regression analyses showed that such considerations are remarkably important in determining the allocation of aid. While donors often stress the importance of democracy or liberal trade policies in developing countries, Alesina and Dollar found that (other things being equal) colonial ties are much more important: "a non-democratic former colony gets about twice as much aid as a democratic non-colony". Similarly, former colonies that are "closed" (in the trade policy sense) appear to get about twice as much aid from the former colonial power as non-colonies that are open to trade (again, in a *ceteris paribus* sense).[2]

Given this importance of political and strategic determinants, one would expect aid allocations to differ substantially between donors… and indeed they do. However,

1. See e.g. Little (1982), pp. 111-4.

2. See Berthelémy (2004) for more recent evidence on the determinants of aid allocation by bilateral donors.

substantial differences between donors remain even if one controls for political and strategic considerations, such as the French interest in West Africa or the importance of the Middle East to the US. The regression evidence shows, for example, that poverty, democracy and openness of the recipient countries are important determinants of Nordic and US aid, but not of Japanese or French aid. Berthelémy (2004) tests whether aid allocation behaviour pools across donors. He rejects pooling and identifies three clusters of donors: the "altruistic" donors (e.g. Norway), the "selfish" donors (including France and the UK) and an intermediate group (e.g. the US).

Hence, rich countries pursue very different objectives with development aid. Promoting development is not even the most important aim and, while trade policy, poverty and democratisation appear central in donors' policy pronouncements, the revealed preference evidence shows that donors differ widely in the extent to which their aid decisions are actually driven by these factors.

Statistical evidence, such as that in the Alesina-Dollar paper, has hardly percolated through to the public debate on the effectiveness of aid. Indeed, some of the disillusionment with aid would otherwise be hard to understand. Obviously, one can hardly expect aid to be effective in terms of, say, poverty reduction or economic growth, if neither of these are the donor's sole or even principal aim. Amongst experts, this point is known and accepted. Collier and Dollar (2002), for example, stress that their "poverty-efficient aid allocation" implicitly takes global poverty reduction (in the headcount sense) as the sole objective of aid. While that measure is plausible (and in accordance with the first of the Millennium Development Goals), donors obviously also have other objectives. Collier and Dollar emphasise that their calculations make the trade-off explicit: the implication of pursuing other objectives can be calculated as the poverty reduction foregone by deviating from the poverty-efficient allocation. However, such subtleties have, as yet, had little impact on the public perception of effectiveness: aid continues to be assessed as if it is solely given to promote development.

Having said that, we will assume that the donor's objective is indeed to promote development, defined, say, in terms of economic growth or poverty reduction. The question "why give aid?" then amounts to asking whether aid is an appropriate instrument to achieve that objective. This question is much more difficult than the earlier, factual one. To answer it one needs, first, a diagnosis: what is it that prevents

or constrains development in a specific country? Secondly, one has to be able to show that aid, in the sense of financial support to a government, will overcome the problem and is more suitable than other instruments.

Remarkably, much of the aid literature is rather vague regarding both these questions (the Millennium project is a notable exception. See Pangestu and Sachs, 2004). This has not always been the case. For example, in the 1960s and 1970s the two-gap model provided a widely accepted diagnosis for low growth in terms of savings and foreign exchange constraints with a clearly defined role for aid. That analytical framework lost credibility some time ago. Its successors have never achieved the same position. The problem aid is intended to solve often remains implicit.[3] For example, while aid is often justified by pointing to capital market imperfections, much aid (as much as half of World Bank lending according to the Meltzer Report) in fact competes with private finance.

The situation is no better regarding the relative effectiveness of aid and other instruments available to donors. To address this seriously one would, for example, have to take a view on the significance of improved access to the markets of rich countries for poor countries and then establish whether development is better served by aid than by dismantling Northern protectionism. This would involve an assessment of the cost and benefits of alternative donor policy stances. While much has been done on the relative importance of aid and trade for development, little is known on how the two compare in terms of political costs in donor countries. If we cannot compare aid and trade in terms of their costs and benefits (from the perspective of the donor) the question simply cannot be answered. How far we are from such an assessment is strikingly illustrated by the Center for Global Development's (CGD) ranking of donors. In the absence of an evidence-based method for aggregating aid, trade and migration, and other policies of rich countries that affect the poor, the CGD had to use essentially arbitrary weights. Cost-benefit analyses such as the recent Copenhagen Consensus initiative are important attempts to put different initiatives on a comparable footing.

3. Or worse. For example, the Meltzer Report (2000) was notoriously inconsistent, giving contradictory diagnoses in different parts of the text. The Interim Report of the Millennium Project (Pangestu and Sachs, 2004), is a very good example of an attempt to derive a case for aid from an explicit diagnosis, but it is, in that respect, exceptional.

How then can the donor achieve his objectives, if these are not fully shared by the recipient government over which the donor has imperfect control? This question has recently arisen in the political economy literature. However, in the late 1990s, a rather different approach was centre stage. Then, the aid effectiveness literature did not attempt to reach a diagnosis or to trace the effect of aid on outcomes in a principal-agent framework. On the contrary, it treated the relationship between aid and economic growth as an empirical question, the outcome of processes that could conveniently be left in a black box.

Burnside and Dollar (2000) established in what circumstances aid was effective in raising growth. While their key result quickly became controversial,[4] their empirical, reduced-form approach became the accepted standard in the field. Regression analysis is used to estimate the relative importance of various growth determinants (geography, initial income, educational endowments, and so on) including aid. Collier and Dollar (2002) similarly took a reduced-form approach to link aid to growth, taking the quality of a country's policy environment into account. They combined a regression estimate of the effect of aid and policy quality (and various controls) on growth with an estimated relationship between growth and poverty reduction. They were then able to answer the question how a given amount of aid should be allocated over countries if the objective is to maximise the aid-induced reduction in poverty. In this work, the implicit answer to "why give aid?" is that aid will raise growth and thereby reduce poverty, provided the policy environment in the recipient country is not hostile.

This approach has been extraordinarily successful. Indeed, there are few examples of economic research so quickly and comprehensively changing the way practitioners view the world. In this paper, we argue that the aid effectiveness debate is reaching its limits. The controversy surrounding the definition of a "good" policy environment and the large unexplained variance in country performance have made the black box approach less attractive than it first appeared. This has contributed to a revival of the political economy approach in the analysis of aid and to in-depth analysis of the nature of the aid "contract".[5]

4. Cf. Benyon (2004) and, in particular, the influential critique of Hansen and Tarp (2000, 2001). There is general agreement now that aid works better in a good policy environment, but considerable controversy on the regression evidence for this still exists, notably on the robustness of the published results (Roodman, 2004). The micro and case study evidence is more compelling (Gunning, 2001).

5. For example, Adam and O'Connell (1999); Azam and Laffont (2003).

There is a second, perhaps more fundamental way in which the aid effectiveness literature is reaching its limits. Its natural unit of observation is the individual country. It thereby implicitly rules out a role for aid in the provision of global public goods, very much an area in which the need for collective action would seem to justify a donor role. For example, a very large part of the cost of civil war is not borne by the country itself but by its neighbours.

The structure of the paper follows this line of argument. In the next section, we argue that rather than asking, "why give aid?", we should ask "what should donors do?" In Section 3, we very briefly review the aid effectiveness literature. This serves as a background for the next three sections in which new rationales for aid are discussed. Section 4 considers the role of donors from a political economy perspective. Here the focus is on the way a donor affects the incentives for a recipient government to choose growth-friendly policies. In Section 5, the role for aid in the context of global externalities is considered. Section 6 discusses the effect of shocks on growth and the way aid can mitigate this effect. Section 7 concludes.

2. What should donors do?

Rather than focusing on the reasons for giving aid, it is useful to ask "what should donors do?" Donors can affect development in many different ways and giving aid may not be the most appropriate channel. For example, trade and migration policies may well be more powerful than aid and in some cases more effective from the donor's point of view. Similarly, there is growing evidence that donors can help to reduce conflict, not only through aid but also through means such as policies that make it difficult to finance rebellions through illegal diamond exports, for example.

The literature on the "coherence" of donor policies focuses on the extent to which these policies work in opposite directions in their effect on poor countries. Here, the issue is quite different: how different donor policies compare in cost-benefit terms. The debate on the efficiency of the allocation of aid had a major impact precisely because it quantified the effect of policy changes in terms of a single objective. This is only beginning to be extended to policies other than aid (as in the Copenhagen Consensus papers, for example).

There are situations where aid can be effective but where it may be dominated by other options open to donors. Take, for example, the effect of policy reforms on risk ratings. Through recent reforms, a number of countries have established a policy environment which, while still in many ways flawed, is sufficient to induce private investment. However, with the private financial sector continuing to assess the environment on the basis of outdated, pre-reform information, these countries are still classified as risky. The economies concerned are often very small, so that the private sector has little incentive to incur the high cost of acquiring information on the nature and credibility of the reforms. For example, while private investors quickly grasped the significance of reforms in China, they failed to do so in small economies such Mozambique and Uganda even when these countries registered spectacular growth rates. In the case of Uganda, risk ratings improved after the key reforms in the early 1990s, but at such a slow rate that it would have taken a generation to reach even the lower end of the risk ratings of the NICs (New Industrialised Countries) necessary to attract substantial foreign investment.[6] Such countries may therefore fail to attract investment, domestic or foreign, despite far-reaching reforms.[7]

In this context, the underinvestment by private agents in information regarding the policy environment and the type of the government justified a donor role. This rationale is becoming increasingly important. As most of the larger economies have already achieved adequate policy environments, recent reformers will increasingly be small (usually African) economies. In these, the problem of underinvestment in information is likely to be particularly acute. The solution might well involve a role, for donors, e.g. a signalling role for a multilateral institution (possibly the IMF) to inform investors about the nature of the policy regime. For donors to play this role they must, of course, possess superior information and be willing to make this available as a public good. If the donor has its own funds at stake, its signal may be more credible, but this is not necessarily so. Private investors need reassurance regarding macroeconomic stability, taxation and the security of property rights. Hence, if a donor were to vary its aid in response to, say, policy changes in education, it may not provide private investors with

6. Haque *et al.* (2000) give a very interesting analysis of the determinants of risk ratings. Jaspersen *et al.* (2000) show that risk ratings are important determinants of investment levels. Collier and Gunning (1999a, 1999b) make the case for a donor role in signalling.

7. See Rodrik (1995) and Gunning (2001).

a relevant signal. There may well be a case for aid here but in a second-best sense: aid would be justified only if there were no scope for rectifying the informational asymmetry.

As this example illustrates, even in cases where aid is effective, it may not be the best way donors can help developing countries.

3. Aid effectiveness

The aid effectiveness literature focuses on the question whether aid works, not why it should be given. While the two questions are distinct, they are, of course, closely related: many question the rationale for aid on grounds of its alleged ineffectiveness. Conversely, when aid would appear capable of lifting a large number of people out of poverty, many believe that this provides sufficient justification. We argue that effectiveness is a necessary but not sufficient case for aid.

Four strands exist within the literature addressing aid effectiveness. The first, which draws both on theoretical and empirical work, argues that conditionality as practiced by most donors (an offer of aid in exchange for promises of specific policy changes) does not work.[8] The second closely related stream stresses that in many (but not all) cases fungibility undermines the donors' intentions: aid released money for uses other than the project ostensibly being financed. The third uses growth regressions to investigate the determinants of aid effectiveness, such as the quality of the recipient country's policy. Burnside and Dollar (2000) do this by introducing an "aid*(policy quality)" interaction term in a growth regression to investigate whether the effectiveness of aid varies with the quality of the policy environment in the recipient country. This approach has been extended to several other potential determinants of aid effectiveness. Guillaumont and Chauvet (2001) interact aid with a measure of vulnerability to trade shocks, Collier and Hoeffler with a dummy measuring a post-conflict situation, Dalgaard *et al*. (2004) find that the effect of aid depends on climate-related factors. The idea is that the effect of aid on growth is conditional on a wide variety of country characteristics, not just on the policies pursued.

8. See e.g. Collier *et al.* (1997), World Bank (1998).

These three strands came together in the World Bank's (1998) widely read *Assessing Aid* report. The first two strands demolished the paternalistic case for aid, *i.e.* the notion that the donor could influence both government policies and project design and management in the best interests of the recipient country. The third strand, as already noted, remains embroiled in econometric controversy, but the conclusion that aid works better in a good policy environment is widely accepted.[9]

If aid is indeed more effective in a good policy environment and if the objective of donors is to reduce a global poverty measure (e.g. the global headcount), then we need to ask how aid should be allocated "efficiently" over countries.[10] Collier and Dollar (2002) pose this question, derive a poverty-efficient allocation and then investigate how actual allocations deviate from it. This paper launched the fourth strand in the aid effectiveness literature. It followed from earlier work investigating to what extent aid was focused on poor countries, e.g. McGillivray (1989).

The aid effectiveness literature has had an extraordinary impact on donors. Somewhat reluctantly, they accepted the message that conditionality does not work and embraced the Collier-Dollar message that a better focus on poor countries with good policies could vastly improve the effectiveness of aid. The Collier-Dollar paper provided a striking measure of effectiveness: the number of people lifted out of poverty by an extra $1 million of aid. By this measure, there are huge differences between donors, e.g. about 350 or higher for Denmark, Norway, the Netherlands and IDA, just over 200 for the US and 150 for the European Commission.[11] A substantial part of the EC budget is used for assistance in Central and Eastern Europe. Of the top ten recipients of EC aid in 2001/2, only one was a Least Developed Country.[12] Clearly, the scope for improvement is huge. Indeed, Collier and Dollar calculated that the number of people lifted out of poverty annually could be *doubled* simply by improving the allocation of a given amount of aid.

9. Note that this statement is considerably weaker than the original Burnside-Dollar claim that aid works only in such an environment.

10. Technically, the allocation of aid is efficient if the marginal productivity of aid in terms of poverty reduction is the same in all countries. Otherwise, it would be possible to achieve a greater impact on poverty with the same aid budget by reallocating aid from countries with low marginal productivity of aid to countries with high marginal productivity.

11. Dyer *et al.,* (2003), Figure 2.

12. Grimm (2004), p. 3.

The econometric evidence indicated that aid effectiveness depends not only on policy quality, but also on other country characteristics such as vulnerability to trade shocks. The corollary that aid allocations should take such characteristics into account has not yet gained wide acceptance. For example, the allocation formula proposed by DFID[13] depends only on the country's poverty and its CPIA (Dyer *et al.*, 2003). Similarly, Collier and Dollar do not investigate determinants of effectiveness other than policy quality. As shown in an excellent discussion in Benyon (2004), their poverty-efficient aid allocations are robust to changes in poverty lines, poverty measures and in the specification of the basic equation: the resulting allocations are highly correlated to each other. However, they are not highly correlated with the actual allocation of aid.

There is some controversy regarding improvements in aid allocation as a result of the aid effectiveness debate. McGillivray (2003) sees signs of improvement, especially for multilateral donors. He notes a rising trend of aid going to the poorer countries: "donors are giving more emphasis to developmental criteria than previously, including selectivity in manners broadly consistent with [the] C-D [Collier and Dollar (2002)] model, although political factors constrain a fuller embracing of such criteria".

The largest donors indeed appear to have a better focus on poverty. However, the multilaterals are less focused than before and the same is true for the Scandinavian countries and the Netherlands. These donors direct their aid towards poor countries, but this does not always coincide with countries with high levels of poverty. They also find that, in spite of the rhetoric, aid is not better focused on countries with good policies. Remarkably, the Scandinavian countries and the Netherlands, while targeting poor countries, direct the highest share of all donors to countries with *poor* CPIA ratings.[14] The share of aid going to countries with good or very good policies declined after 1998 for all donors except for Norway, the DAC average falling from 55.5% in 1993-8 to 50.1% in 1999-2002. As Nunnenkamp *et al.* point out, the World Bank claim that multilateral donors are better able than bilateral ones to focus aid on good policy countries is belied by the evidence: the average for the multilaterals is actually *lower* than the DAC average.

13. The United Kingdom's Department for International Development.

14. Nunnenkamp *et al.* (2004), Table 4 and p. 25.

We have already noted that the aid effectiveness literature has essentially adopted a black box approach: it has studied effectiveness without investigating through what channels aid affects outcomes. Analytically, this is a serious flaw: the implicit suggestion that aid effectiveness establishes a rationale for aid may be attractive to donors, but is a *non sequitur*: there may be other policies available to the donor that are more effective.

Recall the earlier example of a credit-constrained developing country with a good policy environment in which asymmetric information accounted for the capital market imperfection. Creditors were poorly informed about the country's policies and therefore refused to extend credit when this would have been objectively justified. Aid would be very effective in this situation: the returns would be high, reflecting the credit market imperfection. It is not, however, obvious that a case for aid exists here. If the analysis had moved beyond measuring effectiveness to asking *why* aid is effective in this situation, a different policy conclusion would have been reached. The diagnosis of an informational asymmetry problem would have pointed to the need for an intervention to correct the creditors' inaccurate information.

In this sense, the extraordinary success of the aid effectiveness literature may have detracted from the more fundamental question as to what should be done. It is obviously desirable that aid be focused on environments in which it can be effective. However, there may be much higher returns from other donor actions. The danger is that these are not systematically investigated.

There is another danger in the way donors have assimilated the message from the aid effectiveness literature. Donors now seem to favour an "extension" (McGillivray, 2003) of the selectivity approach. This might involve allocating more aid to countries with pro-poor public expenditure. This appears to be a step back towards *ex ante* conditionality: rather than specifying the impact they want to reward, donors appear to be seeking to re-establish a link between aid and policies.

4. Political economy

Explanations for slow growth in developing countries historically focused on their geographical characteristics, such as access to waterways. Now, there is much more

emphasis on the nature of government: unrepresentative regimes and their limited incentives to promote growth are seen as important explanations for slow growth, notably in Africa.[15] Some of the work in this area explicitly investigates the implications for aid of these regimes. A very interesting example is Adam and O'Connell's (1999) influential article and it is worth considering this paper at some length.

In this article, as in much of the modern political economy literature, the state is not seen as a Platonic guardian. In fact, there is a conflict of interest between the government and the private sector. Private investors allocate their resources between a high-yielding taxed activity and a low-yielding non-taxed activity. An increase in the tax rate reduces the incentive to invest in the high-yielding activity and thereby reduces growth. The government faces a Laffer-curve: if it sets the tax rate too high, it drives investors away into the non-taxed part of the economy, thereby reducing its tax revenue. The government uses the state to transfer resources to a favoured group. Indeed, its objective is to maximise the welfare of that group.

In this model, government expenditure is fixed. This could be seen as the amount the government needs to spend in order to prevent a revolt by that part of the population that does not receive transfers: those who are left out need to be bought off. There are two sources of revenue: aid and taxes. Any excess of revenue over the fixed amount of expenditure is used for transfers.

The conflict of interest is clear. For a given amount of aid, the only way the government can increase transfers to its "friends and relations" is by raising taxes, thereby reducing the non-transfer income of the private sector. If the regime is sufficiently representative (which in this model simply means that the size of the favoured group exceeds a certain threshold), there will be what Adam and O'Connell call a "developmental state". The government will choose not to make any transfers. It has no incentive to do so since if the favoured group is large, then only a relatively small part of the cost of the transfers is borne by others. At some point tax-financed transfers are therefore simply not attractive.

Under these circumstances the economy can grow rapidly, not because this is the government's objective, but because the cost of undermining growth through predation

15. See Easterly (2001); for Africa, Collier and Gunning (1999).

exceeds the benefits in terms of increased welfare of the favoured group. Hence, whether the state is predatory or developmental simply depends on incentives. If the regime's power base is sufficiently narrow, an outcome with low-growth (or even stagnation) is optimal from the government's point of view. This is a cynical, but for many countries, realistic view of government.

What is the effect of aid in this model? In the developmental state, *i.e.* if the regime is sufficiently representative, there is no incentive to use aid for transfers. Government expenditure is given, hence aid will be used entirely for tax cuts. Since taxation represents a distortion, aid-financed tax cuts lead to an increase in growth: aid improves the incentive for the private sector to invest in the high-yielding activity. Taxation is inevitably distortionary and thereby growth-reducing: there is no scope for lump-sum taxation. The case for aid here rests entirely on this cost of taxation: aid enables the government to decrease its reliance on the tax distortions that reduce growth.[16] An increase in aid therefore leads to higher growth through reduced taxation.[17]

It should be noted that no conditionality is needed for this happy outcome: there is no need for coercion, since the government has an incentive to use aid in a growth-promoting way. However, a note of warning is in order: this result breaks down for large amounts of aid. It is easy to see why. Since the cost of taxation is increasing, the return to aid (which reduces this distortion) is initially very high. However, the more aid the country receives, the lower the tax rate and hence the lower the distortion. This implies diminishing returns to aid. Indeed, as Adam and O'Connell show, once aid reaches a critical level, the developmental state will collapse and further increases in aid will be used for transfers rather than tax cuts.

If donors pursue a policy of selectivity (giving aid only to countries deemed to have good policies), then the developmental state might well qualify for (unconditional) aid.[18] The policy environment would be qualified as "good", since the private sector has incentives to invest and the state has not been captured by a narrow elite. The fact that

16. This is very similar to the cost-of-taxation argument in Gunning (2001).

17. This is strictly true only at the margin: a large increase in aid could be partly used for transfers.

18. I am less confident on this score than Adam and O'Connell. Donors might well reverse their positive assessment of the policy environment if they come to realise that, at the margin, aid finances tax cuts.

aid has diminishing returns in this context is fully consistent with models of efficient aid allocation.[19] If one is unhappy about the black box aspect of the aid effectiveness literature, then a political economy model complements the analysis: it indicates not just *that* but also *why* aid is effective in this case. It thereby also reveals the limits of a policy of selectivity, as will be discussed below.

The cost-of-taxation argument provides a strong case for aid. Ironically, many donors would probably reject it out of hand. In the rhetoric of "domestic resource mobilisation", aid is supposed to be accompanied by an increase in tax revenue. This idea does not make economic sense: it is, of course, optimal (at least in a static sense) to *reduce* the tax effort in response to additional aid: when expenditure increases with aid, the marginal benefit of that expenditure will, of course, fall. Clearly, the marginal cost of taxation must then also fall in order to preserve efficiency. Hence, it is optimal for aid to be used partly for increased expenditure and partly for a reduction in taxation.[20] In spite of this simple logic, the idea that tax efforts must increase in tandem with aid is firmly entrenched in the minds of donors.

Now consider how (unconditional) aid would be used if the regime were narrowly based (in the sense that the size of the favoured group does not reach the threshold value). In this case, aid would be fully used to increase transfers.[21] Some of this would be used for investment but the economy would be far below its growth potential. In that sense, aid could be termed ineffective. However, in a principal-agent framework (with the donor as principal and the government as agent), the donor's concern for the general welfare in the recipient country will rule out this outcome. Knowing that aid will achieve little, the donor will not offer any, at least not to the government.

Stepping outside the framework of the Adam-O'Connell model, the donor can, of course, attempt to bypass the government by giving aid through NGOs or by entering

19. Note that if there were no diminishing returns, the poverty-efficient aid allocation would give the entire aid budget to a single country.

20/ In the academic literature, this idea is sometimes justified by an appeal to learning-by-doing in tax collection. The cost of taxation could then fall over time but aid could undermine this learning process. Whatever the merits of this case, it does not seem to be the argument donors have in mind.

21. This reflects some strong assumptions in the model. In a more general model, the use of aid for transfers would be decreasing in the size of the favoured group rather than moving from the 100% to the 0% extreme.

into agreements with service providers, e.g., to ensure the provision of basic health and education services. Here again fungibility can raise its ugly head. Suppose the government had provided some of these services in any case in order to satisfy the participation constraint of the groups not receiving transfers. That is, the government would provide some services to the non-favoured groups because it might otherwise lose power. In this case, when donors move in, they in fact take over this role of buying off the opposition from the government. For the government, this implies that the resources allocated to service provision are released and can be used for transfers.

Unwittingly, the donor then finances transfers despite its strict selectivity policy. However, to the extent that donors aim to achieve more than the government would have managed on their own, this is not a problem. Hence, ironically, donor action makes most sense in the worst-run countries: only a minor part of their work is likely to be wasted through fungibility. A corollary of this argument is that there is no good case for giving both aid to a government and bypassing that government through support for NGOs or aid-financed service agreements. This is worth stressing since some donors prefer NGOs to be particularly active in the very countries they have selected for bilateral aid.

Returning to the Adam-O'Connell model, the no-aid equilibrium resembles the outcome of the prisoners' dilemma, or a non-cooperative equilibrium, in the terminology of game theory. As in the prisoners' dilemma, both parties would be better off (they can realise the "gains from aid") if they could bargain and credibly commit to co-operation. Bargaining would be about the amount of aid and the government's taxation policy. Adam and O'Connell show that there are situations in which conditionality (where aid is conditional on a particular tax policy) would ensure a better outcome than the no-aid equilibrium or even an outcome that is efficient (in the Pareto sense).[22] This additional role for aid requires credible commitment.[23]

The selectivity literature has little consideration for such a role for conditionality. It accepts as an empirical fact that "conditionality does not work". Analytically, the reason

22. See Adam and O'Connell (1999), Proposition 7 and Figures 4 and 5.

23. The possibility of improving on the non-cooperative outcome is, of course, not specific to the Adam-O'Connell model. In the theoretical literature, e.g. Azam and Laffont (2003), Moumouras and Rangazas (2004), the case for conditionality always relies on the "gains from aid".

the selectivity literature is impatient with the potential for conditionality in achieving gains from aid is that it doubts the scope for credible commitment. Without commitment the actions of the two parties become time inconsistent; once the bargaining is concluded parties have no incentive to do what they promised.

A well-documented example is the failure of conditionality to achieve sustained trade liberalisation under structural adjustment. When donors offered temporary aid in exchange for trade liberalisation (as they often did in the context of structural adjustment lending), they could not credibly commit to cutting off future aid if the country were to reverse the liberalisation. In many cases, the recipient country, well-aware that the threat of an aid cut-off was not credible, accepted the aid money and then reversed the liberalisation. The cycle could then start again and there are indeed spectacular examples of the same reform being "sold" over and over again to donors.

The problem is two-fold. First, stopping aid is not in the interest of donor agencies. Indeed, their personnel have strong incentives to maintain or increase aid: their career prospects increase with the size of the lending programme. Secondly, an aid cut-off is politically difficult in donor countries, except in cases of human rights violations. A credible threat would be seen as harsh treatment of a poor country and its rationale would be difficult to explain. This is just a specific instance of the more general point that donors are not very good as "agencies of restraint" because of their understandable reluctance to impose penalties. Not wanting to become "tough cops", they become "sympathetic social workers".[24]

As soon as this limitation is accepted, the role for aid is drastically reduced. This is because one then accepts that the "gains from aid" will not be realised, at least not in an essentially static framework such as that used by Adam and O'Connell. Technically, we are back to the non-cooperative equilibrium of unconditional aid. While the selectivity literature does not, of course, adopt this terminology, this is a sensible way to interpret that literature: there are no commitment devices, the gains from aid cannot therefore be realised and hence the best the donor can do is to offer unconditional aid on a take-it-or-leave-it basis to those countries that are likely to use it in a growth-promoting way.

24. Mourmouras and Rangazas (2004, p. 3), Mussa (2002), Willett (2003).

In crude versions of the case for aid selectivity, it is taken for granted that the donor has no effect on the recipient government's behaviour. The government is free to choose its behaviour rule (giving transfers and tax rates as a function of aid and the amount of expenditure) optimally. Aid does not change that rule: it only affects the economy by relaxing the government's budget constraint. Aid will lead to changes in the tax rate and in the amounts transferred but only through this income effect, not by inducing a change in the government's behaviour. In that sense, the donor is powerless: aid has no incentive effect whatsoever. Aid does not change the government's behaviour, it just gives the government more money to do what it wants.

This is an extreme implication of modelling the relationship between donor and government as a one-shot game. If this is relaxed, the government will take into account that its current actions will affect future aid allocations.[25] It may then change its behaviour, e.g. by trying to reduce poverty, not because it is concerned about poverty *per se* but because it is convinced that the donor will increase future aid in response to success in poverty reduction. What has come to be called *ex post* conditionality, where aid is a function of past outcomes, can in this way have an incentive effect.

There is again a commitment issue: this will work only if the government is convinced that the donor will apply its aid allocation rule. As Azam and Laffont (2003) point out, a successful aid programme can be sustained by the "apparently easily enforceable threat" of cutting off aid when performance targets are not reached. One reason that donors have been reluctant to "punish" recipient governments by enforcing that threat is that aid contracts have been vague, leaving much room for discretion. The Meltzer Report proposed a form of performance-based lending with grants tied to the number of vaccinated children. The idea is to achieve credibility by removing the room for discretion. In practice, European donors have moved further in this direction (Adam and Gunning, 2002; Adam *et al.* 2004). A pre-announced rule might make commitment more effective: a reduction in aid would be seen as the logical outcome of poor performance rather than as cruel punishment of a poor country by a "tough cop". However, the problem will never be entirely resolved, since there will always be debate on the extent to which outcomes were influenced by events beyond the government's control.

25. See Collier *et al.* (1997).

In the stylised world of the Adam-O'Connell model, government expenditure is exogenous. Clearly, this rules out precisely the role donors see for themselves: that aid can contribute to growth not through tax cuts, but by financing increased expenditure on infrastructure, health, education, courts and other complements to private capital formation. However, allowing for this would not change the key point: that a narrowly based regime has no incentive to serve the interests of the bulk of the population.

Letting the government decide on the level and composition of government expenditure does, however, change the analysis in an important way: we now face the question how far the government can go in ignoring the interests of large segments of the population. We have already hinted that government expenditure can be seen as a necessary price for avoiding conflict. In the terminology of contract theory, the government has to ensure that the participation constraint of excluded groups is satisfied. If not, it might lose power through an election, civil war or secession.

This could be interpreted as the government providing public sector services to the non-favoured population in exchange for tax revenue. This political equilibrium can disintegrate in two ways. Either the government fails to satisfy the participation constraint of the governed and then loses power. Alternatively, the government may reduce its provision of services while also reducing the tax burden. Consent would be maintained (or at least loss of power avoided) but the government would effectively ignore the private sector. This could happen if it acquired alternative sources of revenue, e.g. oil wealth.

Resource-rich countries (such as oil economies) have typically performed much worse in terms of economic growth than other countries.[26] A plausible explanation is that an unrepresentative regime trying to serve the interests of a small group can, if it has access to resource wealth (such as oil or diamonds), finance its expenditures without having to rely on taxes. If the regime does not rely on taxpayer consent, it does not need to spend on infrastructure, courts and other growth-promoting public goods. In the political economy literature, in the extreme this is seen as a mechanism for state breakdown (Bates, 2004). On a less extreme level, oil wealth can lead to limited relations between the government and the governed. This would reduce growth if public capital were growth-promoting.

26. O'Connell and Ndulu (2004) provide recent growth regression evidence for Africa.

Aid can have the same growth-reducing effect as resource wealth. For an unrepresentative regime, it may be optimal to reduce both the tax burden and the level of government expenditure (in a way that respects the private sector's participation constraint) in response to aid and, of course, to increase transfers. This sheds new light on the Adam-O'Connell story: there, aid (to the extent it was not used for transfers) increased growth by reducing taxation. Here, taxes will be reduced more but government expenditure will also be cut. Hence, aid has two opposing effects on growth, one positive (through tax reduction), the other negative (through a reduction in public sector provision).

That aid may reduce the size of the public sector may seem counterintuitive. The reason is, of course, that in this account the government values public expenditure only for its contribution to keeping the private sector's welfare at some *minimal* level. It is then conceivable that aid is actually detrimental to growth. The private sector's welfare would be unaffected (by construction, the participation constraint must be satisfied), but there would be a substitution of present for future consumption. Sacrificing growth would be optimal: the reduced availability of public capital makes private investment less attractive.

Kanbur and Tuomala (2003) also investigate the effect of aid on taxation. They adopt the optimal taxation framework pioneered by Mirrlees: the government maximises a social welfare function; it must finance a fixed amount of expenditure either from aid or from taxes and it takes into account how its tax policy affects the incentives of private agents. Private agents choose between leisure and supplying effort, taking the government's tax policy as given. The government sets the tax schedule optimally, given the effect of taxation on private agents' choices between effort and leisure.

It is taken for granted in their paper that the donor is powerless: he can affect outcomes only by relaxing the government's budget constraint, not through conditionality. Hence, aid is unconditional and the government responds to aid by recalculating the optimal tax schedule. The leisure-effort choice is, of course, affected; this is an aspect of aid that is usually ignored. The authors use simulation experiments to derive the effects of aid and find substantial incentive effects. Aid is passed on to the private sector through tax cuts and a substantial part (up to half) of the effect of aid is in the form of increased leisure rather than extra consumption. The simulation experiments suggest that aid leads

to more progressive tax schedules. Specifically, it is *optimal* to lower tax rates on the poor disproportionately, given the social welfare function the government aims to maximise. In this sense, aid is shown to be surprisingly powerful in reaching the poor.

Increased tax progressivity induced by aid is a novel argument for aid. However, Kanbur and Tuomala use a very peculiar welfare function. People are ranked by their productivity, and utility weights are the population shares of each productivity class. Clearly, this imposes that the government be concerned about the poor. In practice, governments might attach much lower weights to the welfare of the poor than their share in the population. Indeed, that weight could be zero (as in the Adam-O'Connell framework), if the poor did not belong to the favoured group. In this case, the progressivity result might be overturned: it could be optimal for the government to make tax schedules *less* progressive in response to aid.

The optimal taxation framework is also useful for addressing the question of how aid should be allocated over countries. Kanbur and Tuomala use simulation experiments to see how the optimal allocation of aid varies with characteristics of the recipient countries. They find that if two countries have the same distribution of abilities over individuals but with different means, the poorer country should get more aid, as one would expect. They also find that if the two countries have the same mean but different distributions, the one with the more unequal distribution should get less aid. This result is by no means obvious. Greater inequality (for a given mean) implies more poverty and hence a case for more aid. However, in the more unequal country, optimal marginal tax rates are higher for the poor. This means that when aid enables the government to lower tax rates, less of that relief reaches the poor. Hence, given the government's optimal changes in its taxation policy, aid is less effective in reaching the poor. In the simulation experiments, the latter effect dominates so that the more unequal country should indeed get less aid.

This example nicely illustrates the inevitable limitations of efficient aid allocation calculations. In the Collier-Dollar approach, the less equal country would have qualified for *more* aid. The progressivity of the tax schedule might have led to a worse score for the quality of its policies, but that would have carried too little weight to make a difference. Efficient aid allocations are therefore fragile, much more so than has been stressed in the literature.

5. Externalities

In recent years there has been much discussion on whether aid should finance "international public goods".[27] Some of the arrangements proposed as international public goods are viewed with some suspicion by developing countries. They fear that greater harm than good will derive from them as, for example, in the case of the enforcement of intellectual property rights. However, cross-border externalities are prominent in most diagnoses of stagnation in developing countries. *Prima facie* there is a case for donor involvement, possibly for aid. A DFID paper notes that the cost of reducing poverty may be substantially lower for agricultural research or reforms in the international trading system than for aid in the traditional sense of the word.[28]

Perhaps the most important international public good is peace. Whether aid promotes peace or threatens it has become an important topic within the literature on conflict. One approach sees conflict as rent-seeking. The government is in control of the rents generated by the state and there is a rebel movement attempting to take over. Here aid increases the state's resources. With more at stake, the incentive for the rebels to start a civil war increases, as does the government's incentive to defend itself. However, the government can use the aid to strengthen its defence, raising the cost to the rebels of capturing the state. Aid can therefore deter a rebellion. The net effect is ambiguous. Collier and Hoeffler (2000) stress that aid enables the government to reduce its reliance on exports that generate the rents that attract conflict, such as oil or diamonds. Also, if aid is growth-enhancing it will make conflict less likely by raising the wage of soldiers.[29] Arcand and Chauvet (2001) find evidence that, of these various channels, the first is the most significant: aid helps to reduce conflict because (thanks to fungibility) it enables the government to increase its military spending: *si vis pacem para bellum*. This is another instance of aid being effective in roles which donors typically find unattractive.

Collier (2004) emphasises that many of the stagnating developing countries appear stuck in a conflict trap. Poverty is both a cause and consequence of civil war and having

27. Kanbur (2001) is an excellent introduction into some of the conceptual issues involved.

28. Dyer *et al.* (2002), p. 32.

29. Arcand and Chauvet (2001) argue that aid must be destabilising in this context but do not consider its effect on the cost of conflict.

experienced a civil war greatly increases the probability of a new one. Conflict perpetuates poverty, which itself feeds a cycle of violence. The work of Collier and his associates is highly pertinent to the issue of donor options. First, it should be recognised that the economic cost of conflict is huge: Collier estimates this at some $50 billion. With two wars breaking out annually on average, their cost is double the annual aid budget. Secondly, civil war involves a very large externality. More than half the cost of a civil war is not borne by the country itself but by its neighbours. This justifies international public action. Thirdly, aid to countries with good policies and governance is effective, not only in delivering short-run growth (the effect picked up in the aid effectiveness regressions), but also in reducing the cost of conflict. That effect is strong: Collier estimates that this security benefit alone amounts to 40% of the cost of aid. Fourthly, aid is particularly effective in post-conflict situations, but not immediately following the conflict (when, in fact, donors tend to rush in, as in Afghanistan). It is significantly more effective a few years later (when in practice donors have often turned away). Fifthly, what the conflict literature identifies as effective donor action is usually *not* aid. An example is international collective action to squeeze rebel financing by making it very difficult to sell diamonds (for example) without acceptable proof of origin. This can be very effective: a 10% reduction in the price received reduces the probability of conflict by 12%.[30] Establishing international norms for the use of natural resource rents (such as the Extractive Industries Transparency Initiative) may contribute to reducing the risk of conflict associated with ownership of natural resources. The presence of a UN-sanctioned external military force (as in Sierra Leone) to prevent a relapse into violence can be extremely effective.

A second sector of international public goods concerns health. We will be very brief on this, since it is the topic of another paper at this conference. Malaria and HIV/AIDS involve very large externalities, both domestic and international. The case for donor action in research (such as the Vaccine Purchase Fund) is well known. Treatment and prevention, however, are still far beyond the means of ministries of health in Africa. Current donor initiatives such as the Global Fund are successful but only on a very small scale: the number of treated patients amounts to a few percent of those newly infected. Here, there is a very strong case for aid: the benefits are enormous and governments are unable to borrow against future benefits.

30. See Collier, Hoeffler and Söderbom (2004).

6. Growth, risk and insurance

For several decades now, risk has been a prominent topic in the development literature. Part of this is descriptive, documenting the nature and the magnitude of the risks to which households and firms in developing countries are exposed. These risks include price volatility in world and domestic markets, diseases (notably malaria and HIV/AIDS), unreliable rainfall, poor contract enforcement, irregular availability of inputs or of consumer goods and the risk of conflict such as civil war. That in developing countries economic decisions are taken in a very risky environment is now well recognised.

A more analytic literature, both economic and anthropological, has attempted to interpret a number of institutions (e.g. share cropping, bridewealth, agricultural diversification, trading in small networks, contingent debt repayment) as responses to risk in an environment where formal insurance and credit markets are poorly developed. This literature often makes the point that these responses are costly in terms of either the level or the growth of income. In effect, there is a trade-off between security and income (or income growth). A spectacular example is that when property rights are insecure, agents may *choose* to remain poor: there is no incentive to grow because there is a high probability that wealth accumulation will attract predation.[31]

That there is such a trade-off amounts, of course, to declaring no more than that there is no such thing as a free lunch. Insecurity can be reduced but only at a price: the substitute arrangements for insurance involve an implicit insurance premium. In terms of policy, we need to question whether this implicit premium is unnecessarily high. This premium may exist because risk-pooling inherently involves economies of scale. Informal institutions typically solve adverse selection and moral hazard problems by keeping the number of participants small so as to eliminate asymmetric information: in small groups, one can rely on direct observability.[32] However, risk-pooling in small groups is, of course, costly. Institutional development often takes the form of information sharing and secondary enforcement to remove the necessity of direct observation and to enable large-scale operations. Credit card companies work in this way.

31. Greif and Bates (1995).

32. Udry (1993), in a famous example, documents this for credit transactions amongst rural households in Nigeria. Bigsten *et al.* (2000) analyse it for transactions within small networks of manufacturing firms.

The recent literature on risk adopts this perspective. The central questions then are: would growth be higher if agents had access to insurance or credit markets? If so, by how much? Most importantly, does this involve a case for aid?

In the literature investigating poor African growth, risk has long been a key suspect. Collier and Gunning (1999) stress that there are inherent reasons Africa is more risk-exposed. Many African countries face a combination of semi-arid conditions and a very short growing season. In addition, policy regimes in Africa have often left a legacy of poor infrastructure, poor contract enforcement, weak financial sectors, policy uncertainty and risk of conflict. Governments have thereby not only increased the risk to which private agents are exposed, but they have also undermined risk-coping mechanisms.

Africa's soil characteristics have resulted in very low population densities on much of the continent, in marked contrast to the Indian subcontinent. This makes risk-coping vastly more difficult. Credit markets, informal insurance and other institutions that help agents pool idiosyncratic risks typically require physical proximity to overcome moral hazard and informational asymmetry problems. The RPED evidence showed that an important risk-coping strategy of urban firms is to restrict their transactions to a small number of other firms. This restricts competition and the transmission of information on new markets or technologies. Again, this may well reduce growth.

Basing credit on collateral rather than reputation is seldom feasible. Of the key rural assets, cattle are vulnerable to sickness and theft and land is, in many areas, abundant. Where risks are covariant (most climatic risks and some of the disease risks), low population density precludes insurance. Risk-pooling would work only if carried out over large geographical areas and, with low population densities, it is likely to be undermined by informational asymmetries. This, however, is not inevitable. A combination of local, informal insurance (which could overcome informational asymmetries) and formal reinsurance (to overcome covariance by pooling risk over large geographical areas) might well be feasible but has not yet been attempted.

In addition, in Africa there is a long history of government intervention undermining financial intermediation, particularly in rural areas. For example, control regimes that were in place in the 1970s and 1980s typically involved interventions to channel investment resources to urban activities.

In these circumstances, there is no scope for risk-pooling (through insurance or credit) for covariant risks. Rather, farm households adopt individual risk-coping strategies, notably diversification and consumption smoothing. There is, for example, evidence that in areas where rainfall is more reliable, households are less diversified.[33]

Diversification is, of course, costly: the household foregoes the gains from specialisation. This is a level effect that does not necessarily affect growth. However, level effects easily translate into growth rate effects, e.g. when there are indivisibilities in investment and imperfections in credit markets.

The second risk-coping strategy common in the rural economy is consumption-smoothing, typically using food stores, livestock or both. These assets are themselves subject to substantial risks (livestock illnesses, theft, vermin and spoilage). Hence the consumption-smoothing strategy can be very costly.

Suppose the scope for risk coping is indeed very limited. What does that imply for the effect of risk on growth? Risk can affect growth either *ex post* or *ex ante*. In the *ex post* case, a household or firm changes its investment (up or down) once it is hit by a positive or negative shock. The agent's perception of risk is of no importance: growth is affected only if and when shocks actually occur. Intuitively, such effects cancel out: one would expect that for a sufficiently long series of shocks there is no net effect. Investment would be higher after a positive shock, lower after a negative one and, on average, one could expect the growth path to be the same as in the absence of risk. However, this is true only in very special cases. In general, positive and negative *ex post* effects do not cancel each other out and the long-run effect on growth can be substantial.

However, risk can affect growth even if no shocks occur. This is the *ex ante* effect of risk. Here, the question is whether the household or firm changes its behaviour when it perceives a change in risk, *i.e.* by investing less for a *given* value of the variable of interest, say rainfall for the household or the price of its output for the firm. In this case, the effect of risk will be to alter the following year's income, controlling for the current level of income and also for the current shock. Risk then affects growth through a change

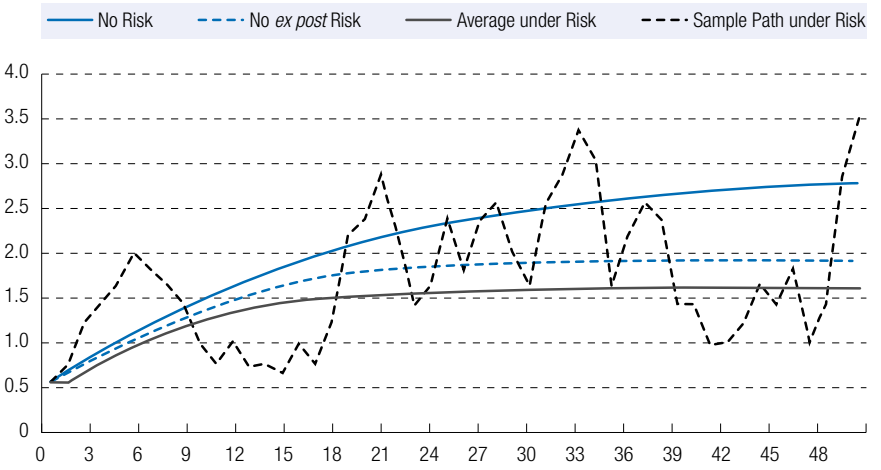
33. See the discussion in Collier and Gunning (1999) and the sources cited there.

in investment behaviour rather than through the realised shocks. To illustrate, when a civil war breaks out, investment will certainly be reduced—the *ex post* effect. However, if agents perceive a risk of civil war, investment may be very much reduced even in long periods of peace. This is the *ex ante* effect.

Analytical work suggests that this effect can also be substantial. Unfortunately, there is very little empirical material to use as a basis for investigation. The effect of risk has typically been studied by using a measure of shocks (e.g. an illness in the family) in a regression explaining income, consumption or investment. This at best identifies the *ex post* effect, but not the *ex ante* effect. The regression indicates the quantitative importance of a shock, e.g. by how much household expenditure will be reduced if a working household member falls ill. It cannot, however, indicate household expenditure in the absence of risk, *i.e.* if the household knew that there would be no future illnesses.

One of the very rare estimates available is Elbers *et al.* (2003). This study of smallholder households in Zimbabwe used panel data for a 20-year period. The authors estimated a model of household behaviour under risk and then used that model in simulation experiments to assess how risk affects growth.

Figure 1.
Accumulation and Risk in Zimbabwe



Source: Elbers *et al.* (2003)

The key results, for a particular household, are summarised in Figure 1. This shows the real value of the household's capital stock (basically livestock) over a 50-year period, starting at *t = 0* from the household's actual starting position in 1980. The figure shows four paths. The "sample path under risk" is one possible growth path, defined by a particular series of fifty randomly drawn shocks, one for each year. One gets a good impression of how much risk this household faces by noting the volatility of this path: its herd size frequently changes by as much as 50% in the course of just one or two years.

A large number (100,000) of such growth paths were generated. In each year, the expected value of the household's capital stock was then calculated as the mean over these paths. The time path of this mean is shown as "average under risk". The averaging procedure obviously removes the volatility of the growth path. This path shows the herd the household would *expect* to attain at future dates, from the starting point of *t = 0*. The path shows very rapid growth in the first 20 years, before levelling off.

The path "no *ex post* risk" shows what would have happened if the household correctly perceived the risk it faced but it never actually received any shocks. In other words, the household makes its investment decisions knowing that in the following year it faces risk, characterised by a given distribution of shocks. As it happens, in the simulation these shocks (positive or negative) do not occur. Obviously, this simulation setup is artificial by design: it enables us to isolate the *ex ante* effect of risk. Finally, if the agent experiences no shocks and is also fully aware that he faces no risk, one obtains the "no risk" path.

The effect of risk on the level of the capital stock can now be measured in the figure as the vertical distance between the "no risk" and "average under risk" paths. Clearly, the effect is massive: the household would have accumulated much more capital in the absence of risk. The effect can be broken down into the *ex ante* and the *ex post* effect by measuring the distance from the "no risk" to the "no *ex post* risk" path and from there to the "average under risk" path. Note that for this household, the total effect of risk is dominated by the *ex ante* effect.

These results also apply to the sample households as a group. In the sample, risk reduces the expected long-run value of the capital stock 46% below the steady state

value in the deterministic case. This is a striking result. Risk not only makes growth highly volatile (illustrated vividly by the "sample path under risk"), but also significantly lowers growth *on average*, so much so that (in expectation) the capital stock only reaches half the value it would attain in the absence of risk. Of this reduction, two-thirds is accounted for by the *ex ante* effect, and the residual by the *ex post* effect.

This also is a remarkable result. In much of the empirical literature (e.g. on vulnerability), it is implicitly assumed that the actual shocks are an adequate measure of the effect of risk. This is not so, however. Most of the impact is internalised in the form of different investment decisions. Hence, growth is significantly affected *even when no shocks occur*.

The Zimbabwe study appears to be the first micro-based study to find a strong effect of risk on growth. Rosenzweig and Wolpin's (1993) well-known paper uses Indian ICRISAT data and concludes that introducing formal insurance has no effect on growth. Their explanation for this is that households engage in efficient informal insurance (risk-pooling within the village), so that if actuarially fair insurance were to be offered, there would be no takers. Some technical problems with the way they model risk exist however. If we accept their finding, though, it can obviously only be true if risk is largely idiosyncratic. Otherwise, there would be only limited scope for informal insurance.

Whatever the case in the ICRISAT villages, many of the risks in developing countries are covariates, related to climate, warfare and government policies. In Africa, covariant risks are dominant and typically no formal financial markets exist to enable risk-pooling at a higher level or borrowing at a given interest rate. Under these circumstances, the situation will resemble that in the Zimbabwean villages with the key risk-coping strategy being self-insurance through consumption-smoothing. This renders the Zimbabwe results interesting. However, whether the key result—that risk in these circumstances massively reduces growth—generalises to other contexts is unknown. Clearly, this is a key priority for applied research.

Suppose the Zimbabwe results do generalise. What then are the policy implications? A substantial part of the risk faced by rural households is covariant. For example, the largest shock the Zimbabwean smallholders experienced was the widespread drought of 1991/2. Traditional types of informal insurance, relying on local observability and,

hence, small-scale risk-pooling, are obviously powerless in such situations: risk-pooling breaks down when all participants in the scheme are hit simultaneously.

Reinsurance is the only solution to the problem of covariance: a risk that is covariant at, say, the national level becomes idiosyncratic at a higher level. Risk-pooling will work if carried out at a sufficiently high level, so that the risk of a drought in Zimbabwe is included in an insurance portfolio with other risks, such as frost in Brazil, which are weakly correlated. In the case of climatic risks such as droughts, the problems of asymmetric information (and hence adverse selection and moral hazard), which often bedevil insurance contracts, do not preclude insurance. Objective (and difficult to manipulate) information on weather outcomes is reliably collected by weather stations. It is therefore feasible to write insurance contracts in terms of outcomes such as rainfall. Nevertheless, such contracts are rare. The most likely explanation is that whoever is locally trusted has no access to international reinsurance markets. Conversely, international insurers hold no position of trust in village economies. Their offer to provide insurance in return for a pre-paid premium would simply not be credible.

The solution probably lies in linking local institutions, e.g. ROSCAs or other forms of microfinance, which have acquired trust and are able to enforce compliance, with formal insurance companies that can offer reinsurance. Can this be left to the market? Probably not: such schemes do not yet exist and are only beginning to be discussed. A major hurdle is the perception that poor people cannot afford to pay for insurance. This is a major misconception. As the Zimbabwe results vividly illustrate, people may respond to risk in ways that reduce their expected income very substantially. It follows that they would be willing and capable of paying a substantial percentage of their income as an insurance premium. The resulting market failure calls for public action. This could take the form of a public-private partnership whereby a donor temporarily shares the reinsurance risk. This might well be a donor role with a very high return.

Thus far, we have discussed risk at the micro level. In the literature on trade shocks, the focus is on the impact of shocks at the macro level. Here too, there might be a case for donor action. Guillaumont and Chauvet (2001) present evidence that aid is particularly effective in economies vulnerable to trade shocks. This amounts to a critique of the Collier-Dollar efficient aid allocation: the allocation cannot be based solely on a country's poverty and the quality of its policies. Initially, this point did not receive the attention it

merited: the paper was relatively neglected, while econometricians engaged in a ferocious and acrimonious debate about the robustness of the Burnside-Dollar results.

The concept of vulnerability which Guillaumont and Chauvet used was problematic in the sense that it did not distinguish between the effect of risk *ex ante* and *ex post*.[34] This is significant since the policy implications are very different. On the one hand, shock-prone countries should get more aid (to compensate for the *ex ante* effect) while, on the other, there is a case for insurance, compensating for shocks after they occur.

Dehn (2000) in his analysis of the effect of risk on growth draws this distinction very carefully in his empirical work. He estimates a forecasting equation for a country's export prices and then derives the distribution of the forecasting errors. The variance of this distribution is then used as a measure of the risk the country faces in that particular market. Outcomes in the 5%-tails of the forecasting error distribution are taken to be shocks. Hence, a price change is treated as a shock if it is considered as a surprise, in the sense that it is an outlier in the distribution of forecasting errors.[35] Dehn found that trade shocks matter, but not in the *ex ante*, vulnerability sense. However, negative shocks do reduce growth.

The distinction is incorporated into an aid effectiveness regression by Collier and Dehn (2002). They augment the standard Burnside-Dollar regression with shocks measures and a "shocks*(increase in aid)" interaction term. This term is significant: aid is particularly effective in the wake of a negative shock. This can be seen as a special case of the capital market imperfection case for aid: growth is reduced by the country's inability to borrow. The EU formerly operated an insurance programme, Stabex, compensating developing countries for export price shocks.[36] The IMF is considering strengthening such forms of country insurance.

34. Gunning (2001).

35. This is problematic in two ways. First, it introduces a discontinuity that seems difficult to justify. A very large price change would not be treated as a shock for a sufficiently high variance of the error distribution. Secondly, a point made by Roodman, the same price change could be classified differently by two countries that differed in their assessments of the risk they faced. Dehn's measure is appropriate if one seeks to measure the extent to which an outcome is surprising, but not when the purpose is to assess the *ex post* effect of risk. For this, as Dehn's thesis supervisor, I take full responsibility.

36. See Collier *et al.* (1999) for a critical analysis of the Stabex programme.

The empirical work in this area has not yet reached consensus. In Roodman's very careful robustness analysis of the aid effectiveness literature, the Guillaumont-Chauvet results appear to be driven by outliers, while the Collier-Dehn results are fairly resilient. However, since Collier-Dehn use the change in aid rather than its level in the interaction term, their results are not directly relevant for aid allocation. In addition, Dehn's decomposition does not correspond exactly to the distinction between *ex ante* and *ex post* effects of risk. However, suppose that both studies are right: that aid is particularly effective in shock-prone economies and that effectiveness arises because shocks affect growth in the *ex post* sense. Then that would establish not a case for aid (as Guillaumont-Chauvet argued), but a case for insurance. There is a powerful argument for a donor role, but not for aid.[37]

Cordella and Yeyatti (2004) investigate how country insurance affects domestic reforms. In their model, when deciding on a programme of reforms, the government takes into account that, if a crisis (caused by exogenous volatility) occurs, the benefits of the reforms will be reduced. Under insurance, there is therefore a greater probability that reform will be undertaken. This mechanism might provide an additional benefit of country insurance for trade shocks.

7. Conclusion

After half a century of development assistance, the case for aid is still far from clear. To the extent that donors give financial aid to developing country governments in order to promote welfare in these countries (and much aid is, of course, given for very different reasons), it is often neither clear what the diagnosis of the poverty problem is, nor whether aid is the best medicine.

Issues surrounding the problems aid is supposed to address and whether aid has a comparative advantage over other policies, have been pushed to the periphery in the excitement of the debate over aid effectiveness. That aid in certain circumstances can lift a large number of people out of poverty appears to be taken as a sufficient answer

37. This is, of course, just another illustration that, for there to be a case for aid, effectiveness is necessary, but not sufficient.

to the "why give aid?" question. Such results are a necessary justification, but are not on their own sufficient.

Credit market imperfections were once considered a self-evident justification for aid. If developing countries could not borrow in spite of the high potential returns (reflecting their capital scarcity), then their growth would be constrained. The case for aid was simply that it allowed countries to overcome this market imperfection.

Many developing countries currently continue to receive aid despite being able to borrow in international capital markets. Conversely, if countries have no access to capital markets, it is often for reasons that would make aid ineffective. Not surprisingly, the critique from the Right is often that a country either does not need aid, or that it would waste it.

The assumption that a country with a set of sufficient policies to make aid effective would *ipso facto* have access to capital markets is, however, very unrealistic. Country risk ratings lag very seriously behind changes in policies. Hence, a country with a very good policy environment may still find itself excluded from foreign borrowing for quite a long period. This diagnosis justifies donor action, although not necessarily aid. The best solution would be for the informational asymmetry problem—the government knows its own risk type but creditors do not—to be addressed directly. This could be effected by giving a specialised donor agency such as the IMF (with better information about the country than private investors) a signalling role. So long as this has not been achieved, the credit market imperfection resulting from informational lags provides a second-best case for aid. As demonstrated in the aid effectiveness literature, the impact of aid in these economies can be very high. Effectiveness is highest if the donor's commitment to selectivity is credible. Donors should be willing to be "tough cops" rather than "sympathetic social workers".

The second rationale, which is related to the first, derives from the cost of taxation. If there is a constraint on foreign borrowing, this will be reflected in domestic interest rates so that domestic borrowing is very costly. In these circumstances, governments will either have to economise on public sector provision or raise taxes. Aid can then be growth-promoting, either by enabling tax cuts, making private investment more attractive, or by increased public spending to achieve an enabling environment for private

investment. Here, the case for aid relies on the combination of a capital market imperfection and the cost of taxation.

It has long been argued that, through *ex ante* conditionality, donors can improve the effectiveness of a government's spending. Understandably, donors like to see themselves in this light, but this rationale has been demolished by the evidence in the aid effectiveness literature, both micro and macro. This empirical evidence should not come as a surprise. There are excellent theoretical reasons why *ex ante* conditionality is desirable: it amounts to a contract that is Pareto superior to the non-cooperative outcome in which the donor either refuses to give aid or extends it unconditionally. However, donors have been unable to find the commitment devices needed to realise the gains available under *ex ante* conditionality. Without a commitment device (e.g. a credible threat to cut off aid), conditionality breaks down. Under selectivity, donors extend aid on a take-it-or-leave-it basis (presumably with either the cost of taxation or credit market imperfection rationale in mind), but do not attempt to influence the use made of their aid.

The third rationale is that aid can have an incentive effect, even under selectivity. For this to work, the rule whereby the donor links aid entitlement to past performance must be both transparent and credible. In that sense, there still is a commitment issue. The search for performance-based conditionality can be seen in this light. The European Union has developed initiatives for linking its aid to performance.[38]

The fourth rationale for aid is the financing of international public goods, notably those related to health. Here, the case rests on the difficulty of borrowing to finance the provision of these goods.

We have argued that in two areas, conflict and risk, there is a strong case for donor action (support for reinsurance of climatic risks, insurance for trade shocks in the form of compensation schemes, criminalising the financing of rebel activity, international rules for resource rents, external military involvement in conflict situations), but that this need not take the form of aid.

38. At the same time, however, donors move in the opposite direction, "extending" selectivity to reward pro-poor expenditure, for example. This extension is likely to undermine commitment.

Ironically, this list runs counter to much donor activity. Capital markets may be slow in recognising the creditworthiness of recent reformers, but donors are often reluctant to give aid to these countries on the spurious ground that if your policies are good, you do not need aid. They thereby reinforce the problems faced by these countries in financing growth: where there is a case for aid, it is not well recognised. While aid can be effective following shocks, the Stabex scheme was abandoned. Similarly, while there is a case for aid in terms of the cost of taxation, this involves aid leading to lower "tax effort", which is contrary to donors' intentions. One channel through which aid helps to avoid conflict is through fungibility: aid, in fact, finances defence expenditure that deters conflict. In all of these cases, aid can work, but not in a way donors find particularly attractive.

REFERENCES

Adam, C.S., G. Chambas, P. Guillaumont, S. Guillaumont Jeanneney and J.W. Gunning, (2004), "Performance-based Conditionality: A European Perspective", *World Development*, Vol. 33, pp. 1059-1070.

Adam, C.S. and J.W. Gunning, (2002), "Redesigning the Aid Contract: Donors' Use of Performance Indicators in Uganda", *World Development*, Vol. 30, pp. 2045-2056.

Adam, C.S., and S. O'Connell (1999), "Aid, Taxation and Development in Sub-Saharan Africa", *Economics and Politics*, Vol. 11, pp. 225-254.

Alesina, A. and D. Dollar, "Who Gives Foreign Aid to Whom and Why?", *Journal of Economic Growth*, Vol. 5, pp. 33-63.

Arcand, J.-L. and L. Chauvet (2001), "Foreign Aid, Rent-Seeking Behaviour and Civil War", mimeo, CERDI, Université d'Auvergne.

Azam, J.-P. and J.-J. Laffont (2003), "Contracting for Aid", *Journal of Development Economics*, Vol. 70, pp. 25-58.

Berthelémy, J.-C. (2004), "Bilateral Donors' Interest vs. Recipients' Development Motives in Aid Allocation: Do All Donors Behave the Same?", paper presented at the HWWA conference on The Political Economy of Aid, Hamburg, December 9-11.

Beynon, J. (2003), "Poverty-Efficient Aid Allocations – Collier/Dollar Revisited", ESAU Working Paper 2, ODI, London.

Bates, R.H. (2004), "Political Violence", paper for the AERC project "Explaining African Economic Growth", mimeo, Harvard University.

Bauer, P.T. (1984), *Reality and Rhetoric: Studies in the Economics of Development*, Weidenfeld and Nicholson, London.

Bigsten, A., P. Collier, S. Dercon, M. Fafchamps, B. Gauthier, J.W. Gunning, A. Oduro, R. Oostendorp, C. Pattillo, M. Söderbom, F. Teal and A. Zeufack (2000), "Contract Flexibility and Conflict Resolution: Evidence from African Manufacturing", *Journal of Development Studies*, Vol. 36, 2000, pp. 1-37.

Burnside, C. and D. Dollar (2000), "Aid, Policies and Growth", *American Economic Review*, Vol. 90, pp. 847-868.

Collier, P. (2004), "Development and Conflict", CSAE, University of Oxford, October.

Collier, P. and J. Dehn (2001), "Aid, Shocks and Growth", World Bank Policy Research Working Paper 2688.

Collier, P. and D. Dollar (2002), "Aid Allocation and Poverty Reduction", *European Economic Review*, Vol. 46, pp. 1475-1500.

Collier, P., P. Guillaumont, S. Guillaumont Jeanneney and J.W. Gunning (1997), "Redesigning Conditionality", *World Development*, Vol. 25, 1997, pp. 1399-1407.

Collier, P., P. Guillaumont, S. Guillaumont Jeanneney and J.W. Gunning (1999), "Reforming Stabex", *The World Economy*, Vol. 22, pp. 669–82.

Collier, P. and J.W. Gunning (1999), "Explaining African Economic Performance", *Journal of Economic Literature*, Vol. 37, pp. 64-111.

Collier, P. and J.W. Gunning (1999a), *Trade Shocks in Developing Countries*, Oxford: Oxford University Press.

Collier, P. and J.W. Gunning (1999b), "The IMF's Role in Structural Adjustment", *Economic Journal*, Vol. 109, pp. F634-651.

Collier, P., A. Hoeffler and M. Söderbom (2004), "On the Duration of Civil War", *Journal of Peace Research.*

Collier, P. and S. O'Connell (2004), "Opportunities, Syndromes and Episodes", paper for the AERC project "Explaining African Economic Growth", mimeo.

Cordella, T. and E. Levy Yeyati (2004), "Country Insurance", IMF Working Paper WP/04/148.

Dalgaard, C.-J., H. Hansen and F. Tarp, "On the Empirics of Foreign Aid and Growth", *Economic Journal*, Vol. 114, pp. F191-216.

Dyer, N., J. Beynon, M. Butler, S. De, P. Landymore, C. Porter, S. Richards, M. Smart, M. Speight and R. Turner (2003), "Strategic Review of Resource Allocation Priorities", Department for International Development Discussion Paper, London.

Dercon, S. (2004), "Risk, Insurance and Poverty: A Review" in S. Dercon (ed.), *Insurance against Poverty*, Oxford: Oxford University Press.

Dehn, J. (2000), "Commodity Price Uncertainty and Shocks: Implications for Investment and Growth", D.Phil. thesis, University of Oxford.

Easterly, W. (2001), *The Elusive Quest for Growth*, Cambridge, Mass.: MIT Press.

Eaton, J. (1989), "Foreign Public Capital Flows" in H. Chenery and T.N. Srinivasan (eds.), *Handbook of Development Economics*, Vol. 2, Amsterdam: North-Holland, pp. 1305-1386.

Elbers, C., J.W. Gunning and B. Kinsey (2003), "Growth and Risk: Methodology and Micro Evidence', Tinbergen Institute Discussion Paper TI 2003-068/2.

Greif, A. and R.H. Bates (1993), "Organising Violence: Wealth, Power, and Government", mimeo, Stanford University.

Grimm, S. (2004), *European Development Cooperation to 2010*, ODI, London.

Guillaumont, P. and L. Chauvet (2001), "Aid and Performance: a Reassessment", *Journal of Development Studies*, Vol. 37, pp. 66-92.

Gunning, J.W. (2001), "Rethinking Aid", in B. Pleskovic and N. Stern (eds.), *Annual World Bank Conference on Development Economics 2000*, World Bank, Washington, pp. 125-144.

Gunning, J.W. (2004), "Shocks, Risk and African Growth", paper for the AERC project "Explaining African Economic Growth".

Haque, N. ul, M. Nelson and D.T. Mathieson (2000), "Rating Africa: The Economic and Political Content of Risk Ratings", in P. Collier and C. Pattillo (eds.), *Investment and Risk in Africa*, Macmillan, London.

Hansen, H. and F. Tarp (2000), "Aid Effectiveness Disputed", *Journal of International Development*, Vol. 12, pp. 375-398.

Hansen, H. and F. Tarp (2001), "Aid and Growth Regressions", *Journal of Development Economics*, Vol. 64, pp. 547-570.

Hoogeveen, H. (2001), *Risk and Insurance in Rural Zimbabwe*, Tinbergen Institute Research Series no. 247.

Jacquet, P. (2003), "Comments on the Summary Paper by Mark McGillivray", Joint OECD-DCD/Development Centre Expert Seminar on "Aid Effectiveness and Selectivity: Integrating Multiple Objectives into Aid Allocations", Paris.

Jaspersen, F., A.H. Aylward and A. D. Knox (2000), "Risk and Private Investment: Africa Compared with Other Developing Areas", in P. Collier and C. Pattillo (eds.), *Investment and Risk in Africa*, Macmillan, London.

Kanbur, R. (2001), "Cross-Border Externalities, International Public Goods and their Implications for Aid Agencies", mimeo, Cornell University.

Kanbur, R. and M. Tuomala (2003), "Incentives, Inequality and the Allocation of Aid when Conditionality Doesn't Work: An Optimal Nonlinear Taxation Approach", mimeo, Cornell University, June.

Little, I.M.D. (1982), Economic Development: Theory, Policy and International Relations, Basic Books, New York.

McGillivray, M. (1989), "The Allocation of Aid Among Developing Countries: A Multi-Donor Analysis Using a Per Capita Aid index", *World Development*, Vol. 17, pp. 561-568.

McGillivray, M. (2003), "Aid Effectiveness and Selectivity: Integrating Multiple Objectives into Aid Allocations", Summary Paper of the OECD-DAC/Development Centre Experts' Seminar.

Meltzer Report (2000), www.house.gov/jec/imf/meltzer.htm.

Mourmouras, A. and P. Rangazas (2004), "Conditional Lending under Altruism", IMF Working Paper WP/04/100.

Mussa, M. (2002), Argentina *and the Fund: From Triumph to Tragedy*, Washington: IIE.

Nunnenkamp, P., G. Canvire and L. Triveno (2004), "Bilateral and Multilateral Aid: Targeted to the Needy and Deserving?" mimeo, Kiel Institute for World Economics, October.

Pangestu, M. and J. Sachs (2004), "Interim Report of Task Force 1 on Poverty and Economic Development", mimeo, Columbia University.

Rodrik, D. (1995), "Why Is There Multi-Lateral Lending?" in M. Bruno and B. Pleskovic (eds.), *Annual World Bank Conference on Development Economics*, World Bank, Washington.

Roodman, D. (2003), "The Anarchy of Numbers: Aid, Development, and Cross-Country Empirics", Center for Global Development Working Paper 32, September.

Rosenzweig, M. and K.I. Wolpin (1993), "Credit Market Constraints, Consumption Smoothing and the Accumulation of Durable Production Assets in Low-Income Countries', *Journal of Political Economy*, Vol. 101, pp. 233-244.

Udry, C. (1994), "Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria", *Review of Economic Studies*, Vol. 61, pp. 495-526.

Willett, T. (2003), "IMF Conditionality and the New Political Economy of Ownership", mimeo, Clermont Graduate University.

World Bank (1998), *Assessing Aid: What Works, What Doesn't and Why*, Oxford University Press, for the World Bank, Oxford.

Comment

by
Patrick Guillaumont
CERDI, Université d'Auvergne

Jan Gunning has given an extremely rich report that approaches a multitude of subjects in an original manner; this commentary cannot attempt to address all of them. My intentions thus will be selective. I would like first of all to draw attention to the elegant manner in which the author, starting from the question of "why give aid?", has led us to another—which he considers, in the end, more important: "what should donors do?" The response to this second question depends, of course, on the objectives pursued, but also, as Gunning stresses, on the cost and effectiveness of means—other than aid—that may be available to the donors to get to the same objectives. Objectives and factors of effectiveness, however, are both subject to debate and the cost of different means—including aid— have never been compared. This I will leave on the sidelines, so as to better tackle objectives and effectiveness. My commentary will thus focus on two main points:

- the relationship between objectives and effectiveness of aid (which raises the question: "why give to this or that country?"). Jan Gunning endorses the idea that aid selectivity has been strongly influenced by the results of research. I would like to examine the question of to what extent this is true, and underline the fragility of the arguments used to judge aid selectivity.
- the relationship between vulnerability (to shocks) and aid effectiveness (or why should aid be given according to vulnerability?). Gunning believes this relationship has been largely ignored and calls for clarification. I will attempt to propose several.[1]

1. At the end of the text, we will indicate references that do not appear in Gunning's paper. We will also refer, on several occasions in this commentary, to the excellent review of literature from Amprou and Chauvet (2004).

1. On selectivity resulting from the aid effectiveness analysis: why give aid to one country rather than another?

Selectivity is the quality of aid allocation between countries. While effective allocation is supposed to reveal the implicit preference of donors, selectivity is supposed to reflect the conformity of effective allocation to those that are "desirable." But desirable according to what? According to objectives and aid effectiveness in contributing to the realisation of these objectives.

Consensus on objectives?

Is there a consensus on aid objectives? It has to do with development, of course, i.e. with growth and poverty reduction in its diverse components, with the increase in abilities—particularly for the most destitute—via health programmes, education, etc. In other words, development, because of its multi-dimensional characteristics, is a multiple objective—as is made clear for that matter in the Millennium Goals. However, the main point of research on aid effectiveness relates to its effectiveness in producing growth. And poverty reduction is generally supposed to result from this in a uniform manner. The plurality of objectives associated with the very idea of development is rarely taken into consideration, with the exception of Kosack (2003). And when it is, objectives are given varying weight. In short, if there is an apparent consensus on effectiveness, it is of a limited scope.

Development itself is not necessarily the only acceptable objective for official development aid or its definition can be more or less broad (security, cultural diversity, etc.). If objectives were identical for all donors and they received the same attention from all of them, the justifications for a multiplicity of donors could be challenged.

Certainly, objectives are not identical. However, the gap between the allocation of their respective aid with regard to a common standard is in general considered, as Gunning reminds us, as a sign of poor allocation and a source of ineffectiveness.

With respect to this, I will allow myself in this regard to express a feeling of discomfort for the dichotomatic—not to say Machiavellian and sometimes naive—interpretation of many (effective) aid allocation models. Here, I refer to models whose variables are

each dressed up with a label that alludes to either the donor's or the recipient's strategy and/or interest, while the majority reflect donor motivation ambiguously: is the preference for small countries (a point we will address later) a sign that donors are looking for votes at the United Nations, or a recognition of these countries' greater vulnerability? Does the preference of certain donors for their ex-colonies translate into their defence of a sphere of influence, or rather an historic obligation of solidarity? Or the feeling, true or false as it may be, that money is more efficiently spent in countries better understood? Does allocation, which successively grows then shrinks (term squared) with per capita income, imply a preference for countries offering prospects, or does it mean that there does exist a limited absorption capacity that grows with income? By using these examples, I do not mean to cast doubt on the usefulness of effective aid allocation models as instruments for revealing donor preference—but rather the usefulness of the interpretations of it that are so often made. Moreover, when these models analyse panel data over 25 years (spanning the period before and after the fall of the Berlin wall) as if donor behaviour had been constant along the way[2], the result is a perplexing political interpretation of the question: why give aid to countries?

Let us nevertheless forget the legitimate plurality of objectives for a few moments (or let us suppose that we are interested only in the allocation and selectivity of the total aid given to developing countries, rather than aid from each source). Let us also consider these goals as being summarised in growth and the reduction of poverty, since the latter is supposed to result from growth.

Consensus on the conditions of effectiveness?

Jan Gunning is right to stress that "the literature on aid effectiveness has had an enormous impact on donors". However, this is only valid to a certain point. Certainly, the author brings up, indeed, the intense debate and numerous critiques that resulted from Burnside and Dollar's initial model.[3] The impact, however, is in line with the initial

2. See, for example, the renowned article on the subject by Alesina and Dollar, 2000; on the other hand, Berthélemy and Tichit (2002) examine the evolution of allocation factors over time.

3. Along with Gunning's references, let us add three others : the recent exchange between Burnside and Dollar (2004) and Easterly, Levine, and Roodman (2004) that occurred in the American Economic Review ; the most recent version of the comparative analysis of Roodman (July 2004) ; and the synthesis of McGillivray (2003).

message—aid is effective only if the policies (or institutions) are sound—rather than with the critiques, complements or nuances that followed.[4]

The debate could have contributed, as Gunning reminds us, to the emergence of a consensus on a softer version of the message, namely that "aid operates better in a good policy environment", instead of the original version, which claimed that aid worked only in such an environment. However, the message of Burnside-Dollar's contributions, *Addressing Aid*, and Collier and Dollar's ABCDs of Aid has been presented often in a poorly nuanced form—one close to the initial message. Thus, as Gunning notices, the supposed consensus, derived from models of effective aid allocation, and selectivity standards—with illustrative properties—totally ignore additions that had been developed in the debate.

It is important to note how much research and debate in the matter has been handicapped by the insuperable obstacles academic researchers have come up against regarding the availability of the World Bank's index of economic policy and institutional quality, the CPIA. The use of this indicator has been limited to the Bank's own studies. We hope that this obstacle will be soon removed.

The most worrying fact is that opinions regarding different donors' aid selectivity tend to be formulated (and established) based solely on Burnside and Dollar's initial message, thus flouting all that has been learned in research since. Notably, the research showing that aid effectiveness (always in terms of growth) depends not only on the policy in place, but also:

- on the economic vulnerability of a country (see infra),
- on the existence of a post-conflict situation (Collier-Hoeffler, 2004),
- on the capacity of aid to allow for an improvement of policies when at first they are wrong (this despite the failures of traditional conditionality brought up opportunely by Gunning),
- on factors more or less precisely grouped under the term of capacity absorption and which condition the rhythm of the reduction of the aid's marginal effectiveness up to the point where it becomes negative, etc.

4. Though it may not necessarily be the most important in places where one expects it to be, as Gunning observes.

A striking illustration of this drift has been provided by the World Bank's *Global Monitoring Report* (2004). This report presents an original analysis of the donors' aid selectivity: for each donor, an indicator of "selectivity" is measured by the sum of absolute values of aid elasticity in relation to the per capita income of the recipients and to the CPIA as they result from the estimation of a function of aid, with (the log of) the population as the only control variable. It is almost secondary to note that (i) the sum of elasticities is made no matter what their level of significance may be (often it is inferior to the most tolerant thresholds), (ii) that aid thus estimated excludes emergency aid (why?) but includes debt cancellations (which results from an international agreement more than a choice of allocation by the donors), and (iii) that the two elasticities are incorporated with equal weight (independent of their supposed impact on growth and poverty reduction, thus referring to an implicit model where aid effectiveness depends exclusively and equally on the levels of per capita income and CPIA). Most important is that this judgement deliberately omits taking into account—as a variable of "selectivity"—characteristics of the countries upon which the effectiveness of aid depends in terms of growth and poverty reduction. The report mentions aid needs that stem from vulnerability or a post-conflict situation, but donors are judged exclusively from the indicator so calculated. *Scripta volent*, *digita manent*[5]… A work in progress in co-operation with AFD examines the difference between the indicator thus found and an indicator based on the same principles (whatever their limits may be) but which also includes aid elasticity in relation to the United Nations' indicator of economic vulnerability (EVI). We will now look at reasons for doing so.

2. On vulnerability, as it influences aid effectiveness, or why give aid to vulnerable countries

In the sixth section of his document, Jan Gunning gives a very stimulating analysis of the influence of risk factors on growth. He makes an important distinction between risk *ex ante* and risk *ex post*, and defends its importance in the analysis of aid effectiveness. To do this, he refers to the work by Collier and Dehn (2001) and to my

5. An even higher level of simplification is reached when donors are compared on a diagram representing on one axis, aid per capita to countries with a good CPIA, and on the other axis, aid per capita to countries with bad CPIA. If the distribution of countries according to their CPIA is normal, we can imagine the sensitivity of the result to the location of the cursor which separates the "good" and the "bad"…

own, which was done in collaboration with Lisa Chauvet (mainly in 1999/2000 but also 2004). These works are based on cross-country economic analyses. The supposed links behind the cross-country regressions appear to be "black boxes"; one must try to understand their contents in order to formulate recommendations relative to aid policy.

A brief return to models where aid effectiveness is a function of vulnerability.

The argument that we initially presented and have since developed (notably Chauvet and Guillaumont, 2004) is that aid effectiveness in terms of economic growth depends more significantly on economic vulnerability than on current economic policy (measured in Burnside-Dollar style)[6]. All things being equal, aid is more effective in vulnerable countries, i.e. aid softens the negative impact that vulnerability has on growth. The supposed reason is that the presence of a high level of aid in countries undergoing shocks (alternating between positive and negative) alleviates the impact of these shocks, whose succession is in itself unfavourable to growth (according to the methods examined, in particular in Guillaumont, Guillaumont Jeanneney and Brun, 1999).

The concept of economic vulnerability used here is a concept of structural vulnerability in which the components are supposed to reflect both the extent of crises undergone by a country and the degree to which countries are exposed to these shocks (but not of their resilience, which depends more directly on economic policy). Thus, in our 2001 article, we included, as a component: instability of exports of goods and services (for external shocks), agricultural production instability (for climatic shocks), and the (log of) population, as the principal structural variable explaining (negatively) the degree of trade openness, and thus exposure to shocks.[7] We should underline here that the narrowness of the demographic base or the smallness of countries is considered as a factor of vulnerability, and it also contributes to the increase of aid effectiveness,

6. Compared significance of the sign of the coefficient of the multiplicative variable (policy * aid) and (aid * vulnerability). An analysis of these works is presented in McGillivray (2003) and Roodman (2004).

7. This indicator is similar but distinct from the EVI, the economic vulnerability indicator used by the United Nations for identifying the LDCs, which includes two/three other components in addition to those which were noted; it is currently under revision. On the other hand, the indicator of vulnerability that we have used also includes the trends in terms of trade. The extension also consisted in having the policy partly endogenous, the improvement of policy depending on its initial level and on the level of aid supplied.

justifying, to a certain extent, no matter what donors interest may be, their preference for small countries.

In subsequent developments of the model (Chauvet and Guillaumont, 2004), we prolonged the analysis of factors contributing to influence aid effectiveness, notably by (i) opposing economic vulnerability to internal political instability—which, despite having a negative additive effect like the former, contributes to a lowering of aid effectiveness instead of increasing it— then (ii) by including political instability of neighbours into the vulnerability, which this time contributes to growth in aid effectiveness (Chauvet 2004).[8]

Type of risk and aid effectiveness

Beyond these tests, we must point out—according to the distinction made by Jan Gunning between *ex ante* risk and *ex post* risk—what differentiates our analysis from that of Collier and Dehn. Collier and Dehn argue that aid effectiveness is stronger during years when the countries undergo negative shocks on the export prices of their main commodities. Their analysis addresses only shocks on prices of commodities and not on all the supposed exogenous shocks with external or climatic origins (and it only takes into account exposure through the rate of exports of the basic products). But it distinguishes shocks (of basic product prices) according to whether they are positive or negative. The positive effect of annual negative shocks multiplied by the *change* in aid (and not an average vulnerability over a certain period multiplied by the level of aid) that serves to test the hypothesis of a larger aid effectiveness in vulnerable countries.[9]

The risk, which is supposed to increase aid effectiveness according to Collier-Dehn, is (as Gunning notes) really an *ex post* risk or real risk (even if the shock is

8. Before coming to the comparison Gunning makes between our work and that of Collier and Dehn, notably in light of the type of risk considered, I'd like to make a clarification on the strength of the results. David Roodman, as Jan Gunning reminds us, has used data given by the authors to put the initial models of Burnside-Dollar, Collier-Dehn, Chauvet-Guillaumont, and also the great critical work of Hansen and Tarp under a "technical control". In the first version (2003) cited by Gunning, we appear to be placed in an intermediate position between Burnside-Dollar (less good) and Collier-Dehn (better). But in the final version (2004) that was published after out correspondence with him, we were "reclassified" ahead of Collier-Dehn while remaining behind Hansen-Tarp… We are awaiting, with confidence, the control of results from our following works, without misjudging the limits of the type of analysis which we used.

9. The authors also have found here a significant effect of positive shocks multiplied by the level of aid.

defined from a forecasting model). It is true that the notion of risk that we are considering could appear more ambiguous, because it is based on pluri-annual instability measurements and may, therefore, pick up an *ex ante* risk at the same time as an *ex post* risk (the *ex ante* or perceived risk depends on previously experienced shocks, i.e. on *ex post* risk). Actually, it has above all to do with an *ex post* risk; the main part of our argument on aid effectiveness in vulnerable countries refers to the capacity of aid to alleviate the consequences of shocks, or *ex post* instability. We should certainly seek to evaluate the aid's capacity to alleviate the consequences of perceived risks; again, for them, it is necessary to dispose of adequate measurements. Such measurements, however, are inevitably open to discussion. This is without a doubt one of the reasons why the impact of perceived instabilities on growth (independent of aid) is found to be less significant than that of *ex post* instabilities (Guillaumont and Deméocq, 1989).

To be honest, research of methods where *ex post* shocks or risks increase aid effectiveness (i.e. aid that alleviates the negative impact of shocks) leads to another distinction between the effect of the average level of aid and that of its contra-cyclic or compensatory character (on which Collier and Dehn focused). The relative influence of shocks on the economy may be alleviated by a *permanent* high level of aid and not only if the aid is transferred in a contra-cyclic manner: if the aid is stable, the flow "aid + exports" is less unstable than exports only. We will propose to shortly test the specific effect of the compensatory character (or, on the contrary, pro-cyclic) of aid, concurrently with the effect of aid conditioned by the average level of vulnerability that we have brought to light.

This distinction is not without implications for aid policy. If aid is more effective when it has a contra-cyclic character (all the more if the greatest effectiveness of aid in the most vulnerable countries is essentially due to such a character), it is suitable to look for aid mechanisms that are truly compensatory— but of which we know the difficulties and failures— or insurance mechanisms supported or guaranteed by aid (see Guillaumont, Guillaumont Jeanneney , Jacquet, Chauvet and Savoye, 2004). If, as we think, the vulnerability increases, the effectiveness of an average level of aid, independent of its contra-cyclic character (probably weak), a permanent priority should be given to aid allocation to the poor and vulnerable countries. In relation to the objective of poverty reduction by growth, vulnerability should thus be one of the criteria of aid selectivity.

We can notice that this is a way to justify the international objective of preferential aid allocation to the LDCs (Least Developed Countries), which are defined by the United Nations according to income and structural vulnerability criteria. The foundation of the LAC category is to identify the poorest countries suffering from structural growth handicaps. Compensating these handicaps is an essential reason for aid.

REFERENCES

Amprou J. and L. Chauvet (2004), "Efficacité et Allocation de l'Aide : Revue des Débats", *Notes et Documents* No.6, Agence Française de Développement.

Burnside C. and D. Dollar (2004), "Aid, Policies and Growth: a Reply", *American Economic Review* 94 (3), pp. 781-784.

Chauvet L. (2004), *L'Aide extérieure permet-elle d'amortir les chocs politiques régionaux ?*, mimeo, CERDI.

Chauvet L. and P. Guillaumont (2004), "Aid and Growth Revisited: Policy, Economic Vulnerability, and Political Instability", in Tungodden, B., Stern, N., Kolstad, I. (eds.), *Towards Pro-Poor Policies – Aid, Institutions and Globalization*, World Bank / Oxford University Press.

Dollar D., and V. Levin, (2004), "The Increasing Selectivity of Foreign Aid, 1984–2002", *Working Paper* 3299, The World Bank, Washington D.C.

Easterly W., R. Levine, and D. Roodman, (2004), "Aid, Policies and Growth: a Comment", *American Economic Review* 94 (3),pp. 774-780.

Guillaumont P. and M. Deméocq, (1989), "Risk and Ratchet Effects of Export Instability", CERDI.

Guillaumont P. (2004), "*On Providing More and Better Aid : More and Better Assessments Needed*", mimeo. CERDI.

Guillaumont P., Guillaumont Jeanneney S., Jacquet P., Chauvet L., Savoye B., (2004), "Attenuating Through Aid the Vulnerability to Price Shocks", in F. Bourguignon and P. Jacquet (eds.) *Economic Integration and Social Responsibility*, World Bank / Oxford University Press.

Guillaumont P., Guillaumont Jeanneney, S., and Brun, J.F. (1999), "How Instability Lowers African Growth", *Journal of African Economies* 8(1), pp. 87-107.

McGillivray M. (2003), "Efficacité de l'aide et Sélectivité : Vers un Concept Elargi", *Revue d'Economie du Développement*, No.4, Décembre, pp.43-62.

Roodman D. (2004),"The Anarchy of Numbers: Aid, Development, and Cross-country Empirics", *Working paper* No 32, Center for Global Development, July.

World Bank (2004), *Global Monitoring Report*.

Comment

by
Jean-David Naudet
Département de la Recherche
Agence Française de Développement

The fundamental error of both the critics and the supporters is to identify constraints and, especially in the excitement of discovering new constraints, over-emphasise their importance, often to the extent of eclipsing other factors relevant to the development process. Perhaps the most amazing aspect of the aid debate over the past thirty years is the way in which apparently upright and sensible academics, analysts and politicians have led the world to believe or wish to believe that the aid relationship can be simply derived and conclusions simply drawn.

Roger C. Riddell 1987

Jan Gunning's paper proposes a broad analytical survey of arguments justifying the case for international aid, concluding by laying down the foundations of international aid justification, i.e. "the case for aid": correcting market and credit imperfections (thus risk management), reducing taxation of productive activities, improving policy-making via incentives which result from a selective aid allocation, and financing global public goods (such as peace, in particular).

I would like to discuss the extent to which the rational use of aid for stimulating growth and reducing poverty—so brilliantly analysed by Jan Gunning—constitutes the case for aid itself. We will thus respond to the question with which his paper is entitled: why give aid?

In summarising the main thrust of Gunning's idea, three questions come to mind: "why give aid?", "what should donors do?", and "how should aid be allocated?"

Gunning concentrates on the second question: "what should donors do?" He argues that the recent economic debate has focused on aid allocation and has now reached its limits, even contributing somewhat to a distancing from the central question of what donors should do beyond simple distribution of aid to recipient countries. He also appears to believe that the question "why give aid?" can be reduced to the second central question of what to do. In effect, if one defines the rational and effective applications of international aid, this justifies simultaneously the aid's *raison d'être.*

I would like to approach these three questions. First, I will make a short remark about the difficulty of settling the overly open question of "what should donors do?" I will then contest the idea that the questions "why give aid?" and "what to do" are equivalent. Finally, I will look at possible ways to expand the debate on aid allocation, and attempt to bring out questions that have been rarely addressed in this heavily debated subject.

1. What should donors do?: an excessively open question

Can the question regarding the best uses of aid be resolved in a general manner? Jan Gunning raises the issue: *"To the extent donors give financial aid to governments of developing countries in order to promote welfare in these countries (and much aid is, of course, given for very different reasons), it is often neither clear what the diagnosis of the poverty problem is, nor whether aid is the best medicine."*

Nevertheless, he proposes and analyses a certain number of relevant uses of aid raised in the introduction: correcting market and credit imperfections, reducing taxation of productive activities, improving policy via incentives which result from selective allocation, and financing global public goods.

Despite the merits of these uses of aid, the reader cannot help asking: why these rather than others? One could also defend—with arguments no less convincing—supporting educational or health systems, or protecting local environments, or even empowering the most vulnerable populations.

Can one really respond in a convincing manner and by a short list to the question of what donors should do in order of priority.

To be more specific, I will formulate the question from two angles.

First, do we have methods objective enough to determine the best uses of aid in absolute terms, in ways where choices are not for the most part determined by the consensus of the moment or preferences of analysts? The constant evolution of development priorities should prompt a certain reserve regarding the capacity to objectively determine the best uses of aid.

Second, can we really elucidate in a holistic manner the question of "what to do" without answering the questions "how?" and "with whom?" The "how?" determines in part the "what?" Because conditionality has been judged poorly effective, direct promotion of reforms has been disappearing little by little from the donors' agenda. Perhaps it is because institutional support has brought disappointing results that it appears lower on the list of priorities. The "with whom?" should also be a central concern in the definition of priorities. One can easily imagine that certain countries have their own ideas of what should be considered high priority—a successful and equitable educational system, for example, or security against climate risk.

The "how?" and "with whom?" determine the nature of the partnership. Jan Gunning regards the question of "what to do?" as having been in part eclipsed over recent years by analyses of effectiveness. This is true, but the issue of partnership has surely received even less attention from economists.

2. Why aid?: motivations and expected effects of aid

In line with the designers of the aid selectivity thesis, Gunning considers that a rationale entirely dedicated to the fight against poverty is a prerequisite to the aid's effectiveness according to the following objective: *"Obviously, one can hardly expect aid to be effective in terms of, say, poverty reduction or economic growth if that was not the donor's sole or even main aim."*

It appears to me, however, that the link between the motivations for an action and its economic effects is much more complex. It is one of the founding questions of classical economics as expressed by Adam Smith in probably one of the most cited passages of economic literature: *"it is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages. Nobody but a beggar chooses to depend chiefly upon the benevolence of his fellow-citizens."*

The self-interest behind the motivation of actors—in our case the donors—does not automatically bring about harmful effects on the community.

The self-interest of actors in their allocation of aid, and in the contents of this aid, has been often illustrated, and generally denounced, notably since the pioneering work of McKinley and Little (1977, 1978a, 1978b, 1979). More recent analyses (Alesina and Dollar, 2000) have also highlighted the strategic, commercial and post-colonial motivations in the granting of aid. The long-dominant view appears to be, in my opinion, not that the effectiveness of earlier aid was not a relevant question, but rather that giving aid for reasons of national interest simply was not seen as conflicting with the beneficial effects of this aid, as long as it was used to finance good projects. In the past, the often-cited example concerning aid effectiveness was the South Korea case (OECD, 1985).

What has changed? Globally, donors' strategic interests have without a doubt lowered in intensity since the end of the Cold War, although they have not disappeared. Furthermore, the reality of a triangular (at least) relation between donors, receiving states and beneficiary populations have progressively replaced the binary donor-recipient scheme. Proponents of selectivity (Burnside and Dollar, 2000) have spectacularly illustrated this new order in asserting that aid was not having any effect in poor political or institutional environments.

The main traditional motivations behind aid, often based on the pre-eminence of bilateral state-to-state relations, were sometimes found to be themselves obstacles to the aid's effectiveness in terms of gains for the beneficiary populations, as Jan Gunning has shown through political economy models. The question of effective aid allocation, and indirectly its motivations, has since become central.

The motivations behind aid are not to be confused necessarily with the development objectives pursued. However, they should be compatible and not conflict with them.

Henceforth, it has been accepted that the central goal of the development community—fighting poverty—should be reflected in overall aid allocation. Yet, the fact that donors' motivations in the allocation of aid are less discernible does not mean that national self-interest has disappeared, and that the motivations of donors have merged with the development objectives determined by the international community.

National strategic interests still exist albeit in new and different forms. International aid is, as in the past, an instrument for promoting values of a universal calling: human rights, democracy, gender equality, equal opportunity. While certain donor motivations have declined, new ones (such as security and environment) have probably appeared. It would be interesting, for example, to analyse how preoccupations about international image are becoming a base motivation behind development aid, as shown by the important position the question acquires during international summits.

Are the new or changing motivations behind aid perfectly compatible with its expected effects in regard to poverty reduction? Proof has yet to be shown. The question of "why give aid?" remains pertinent and cannot be reduced to the demonstration, however necessary, that effective use of this aid exists.

3. How to allocate aid?: skewed debates and real questions

If we stand back and take a look at the recent literature on aid effectiveness and allocation, it is surprising to see just how much importance has been given to this issue. Jan Gunning appears to share this assessment: *"This approach has been extraordinarily successful; indeed, there are few examples of economic research which have so quickly and comprehensively changed the way practical people view the world."* He mentions on several occasions the idea that the recent debate on these questions—summarised by the term selectivity—has henceforth reached its limits and may even be diverting attention from more important questions.

How can such a success be explained? Apparently, not by particularly robust empirical results (see Box 1).

Box 1.
Effectiveness and selectivity: how convincing are cross-country econometric analyses?

Since the mid-nineteen nineties, dozens of cross-country econometric analyses have turned their attention to the extent of general effectiveness of aid (Amprou and Chauvet, 2004). All of these studies have attempted to describe contemporary links in the short term (in general, four years) between macro-economic results (most often growth) and the amount of public aid (ODA) as it is calculated by the DAC[1] or a similar aggregate[2]. The measure of aid by the DAC is an aggregate that includes debt relief, financial aid, technical assistance, humanitarian aid and decentralised co-operation. For example, different operations such as road construction, support of banking system reform, technical and material support of legal systems, and promoting female education are all lumped together by adding up their annual costs. It seems very optimistic, at the very least, to think that we can distinguish the immediate impact, on overall economic growth, of a sum of such disparate expected effects. But in spite of this, it was not until 2004 (to the best of our knowledge) that this difficulty was brought up in an academic publication (Clemens, Radelet, Bhavnani, 2004).

Econometricians of aid effectiveness strive to find significant coefficients for a set of key variables in a growth regression: particularly "aid" and an interactive term "aid multiplied by the quality of institutions". Further interpretation of such growth models—in which *ad hoc* specifications are driven by the search for results for these key variables—would be ill advised. The recent analysis of Burnside and Dollar (2004), which uses an updated data set to confirm the results found in their pioneering article, is a shining example. First off, their model does not bring out a significant effect of the quality of institutions on economic growth. Therefore, their surprising conclusion is that sound institutions are good for growth only if they are associated with aid. Secondly, they emphasise a curious relationship between aid

1. OECD Development Aid Committee.

2. A few authors use Effective Development Assistance (EDA), which considers technical assistance and debt cancellations in a different way from ODA.

and growth: the variable "quality of the institutional environment" takes both positive as well as negative values. For many countries (around sixty), it appears that the impact of the interactive variable "aid * institutional environment" implies that aid has generally a negative impact on growth. Thirdly, it is strange to note that the breakover point of aid's marginal effectiveness coincides opportunely with the "zero"—fixed in an independent and arbitrary manner—of the variable "institutional environment quality".

Even if we overlook the oddness of models that emphasise the selective efficiency of aid, numerous authors have underlined the lack of robustness of their conclusions. The results obtained by Burnside and Dollar (2000) in their most commented paper apparently do not resist, for example, the addition of country fixed effects (Dayton-Johnson and Hoddinott, 2003), (as recommended for panel studies), the extension of the period examined (Easterly, Levine and Roodman, 2003), or minor modifications of the list of countries under consideration (Dalgaard and Hansen, 2001).

To explain the echo met by these works, it is pertinent to evoke the context through which the main idea of aid selectivity has emerged: "aid has more impact when it operates in an favourable institutional environment and must therefore be allocated in part according to this criteria". The appearance of this theme in the World Bank's official discourse dates back to 1993 (Jaycox, 1993). The *Africa in Transition* report of 1995 firmly defended selectivity as one of the principal recommendations. The pioneering study by Burnside and Dollar, which brought scientific support to this thesis in 1997, therefore appeared subsequent to the World Bank's promotion of more selective aid.

The nineteen nineties were marked by a major crisis in development aid, which saw—in the post-Cold War period—declining budgets and more and more questioning of its effectiveness. Aid fatigue seized political authorities, public opinion and even the development community. This was partly the result of admitting that a perverse relationships exists with certain states careless of the well-being of their populations.

The development community needs new motivations and ideas to convince again about the effectiveness of aid, while escaping the trap of automatic aid distribution. Aid selectivity is a founding element of this renewal. It is thus the object of wide promotion, with everything that contributes to reinforcing it being strongly valued. The stir caused by econometric studies establishing selectivity should no longer cause alarm; the

same goes for the fact that among the dozens of researchers having worked on the subject, the only ones who have obtained convincing results in favour of selectivity, to the best of our knowledge, come from the World Bank.

We could have stopped there, with a timely idea that remains in line with common sense and is confirmed by economic analysis: that aid has more impact when it operates in a favourable institutional environment. But such was not the case. Selectivity will launch an intense academic debate and a change in aid institutions (which will be surely marked more in discourse than in practice). The intensity of this debate can be explained without a doubt by the fact that, behind its apparent simplicity, the idea of selectivity hides issues that are fundamental for its advocates and adversaries alike.

For the advocates of selectivity, the question at stake is that of incentives. Jan Gunning clearly shows the critique of conditionality, notably the absence of credible defection threats for donors involved in a programme. Programme conditionality does not constitute an incentive sufficient enough to bring about improvements in policies and institutions in beneficiary countries. Therefore, the issue is to have an aid distribution method that becomes in itself the determining factor of such incentives. [3] Aid selectivity is the construction of an aid allocation procedure that creates an environment of incentives for improving policy and institutions in the poor countries.

For the adversaries of selectivity, the problem appears to be in the conflict between selective aid allocation and the concept a fair aid allocation. What perturbs a good number of analysts is not what is said—which is easily consensual—but what is not said: why would the institutional environment be the only co-factor of aid effectiveness?[4] It is just as intuitive to think that aid can have less impact in countries not only because they are poorly led, but also because they have bigger problems to deal with, such as low literacy rates, or high rates of AIDS. The difference is that such relations are not being tested and are debated even less. If they turn out to be founded, we could easily understand that attributing aid by maximising its effectiveness clashes head on with natural principles of equity.

3. This issue is very close to the *exit versus voice* dilemma, emphasised by Hirschman. If the *voice* has no impact on the beneficiary institutions, the *exit* will.

4. Jan Gunning indeed asks this question*: "Collier et Dollar do not investigate determinants of effectiveness other than policy quality".*

Economists have clashed—with not always convincing econometric fine-tuning—over the explanation of the relationship between aid and growth. Can it be said that the debate has lasted too long and that henceforth it should focus on other questions? Jan Gunning seems to adopt this point of view: *"In this sense the extraordinary success of the aid effectiveness literature may have detracted from the more fundamental question as to what should be done."*

Have the grand issues of aid allocation really been at the heart of the controversies? Hardly any authors have addressed the question of aid allocation and selectivity from a perspective of equity, for example from the point of view of equal opportunity (Cogneau and Naudet, 2004).

Neither have any attempted to answer another, truly delicate question: is the selective allocation system likely to create incentives for a real improvement of institutions?

REFERENCES

Alesina A. and D. Dollar (2000), "Who Gives Foreign Aid to Whom and Why?", *Journal of Economic Growth*, Vol. 5, pp. 33-63.

Amprou J. and L. Chauvet (2004), "Efficacité et allocation de l'aide : revue des débats ?", Notes et Documents N°6, AFD.

Banque mondiale (1995), *A Continent in Transition: Sub-Saharan Africa in the Mid-1990s,* Washington DC.

Burnside C. and D. Dollar (1997), "Aid, Policies and Growth", Policy Research Working Paper N°1777, The World Bank.

Burnside G. and D. Dollar (2000), "Aid, Policies and Growth", *American Economic Review*, 90, pp. 847-868.

Burnside G. and D. Dollar (2004), "Aid Policies and Growth: Revisiting the Evidence", World Bank Policy Research Paper N°3251, The World Bank.

Clemens M., S. Radelet and R. Bhavnani (2004), "Counting Chickens When They Hatch: the Short-Term Effect of Aid on Growth", Center for Global Development, Washington DC.

Cogneau D. and Naudet J.D. (2004), "Qui mérite l'aide ? : égalité des chances, aide internationale et réduction de la pauvreté", Notes et Documents N°7, Agence Française de Développement, Paris.

Dalgaard C-J. and H. Hansen (2001), "On Aid Growth and Good Policies", *Journal of Development Studies,* 37(6), pp. 17-41.

Dalgaard C.J., H. Hansen and F. Tarp (2004), "On the Empirics of Foreign Aid and Growth", mimeo, forthcoming in the *Economic Journal*.

Dayton-Johnson J. and J. Hoddinott (2003), "Aid, Policies, and Growth, Redux", mimeo, Dalhousie University and IFPRI.

Easterly W., R. Levine et D. Roodman (2003), "New Data, New Doubts: Revisiting 'Aid, Policies, and Growth'", Working Paper 26, Center for Global Development, Washington DC.

Hirschman O. Albert (1970), *Défection et prise de parole*, Fayard, Paris.

Jaycox V.K. Edward (1993), "Capacity Building: the Missing Link in African Development", Address to the African-American Institute Conference *African Capacity Building: Effective and Enduring Partnership*", Virginia.

Mac Kinlay R.D. and R. Little (1977), "A Foreign Policy Model of US Bilateral Aid Allocation", *World Politics*, XXX, pp. 58-86.

Mac Kinlay R.D. and R. Little (1978a), "The French Aid Relationship. A Foreign Policy Model of the Distribution of French Bilateral Aid: 1964-1970", *Development and Change*, 9, pp. 459-478.

Mac Kinlay R.D. and R. Little (1978b), "A Foreign Policy Model of the Distribution of British Bilateral Aid: 1960-70", *British Journal of Political Science*, 8, pp. 313-332.

Mac Kinlay R.D. and R. Little (1979), "The U.S. Aid Relationship: a Test of the Recipient Need and Donor Interest Models", *Political Studies*, XXVII, pp. 236-250.

OECD (1985), *Twenty-Five Years of Development Co-operation*, OECD, Paris.

Riddell C. Roger (1987), *Foreign Aid Reconsidered*, ODI/John Hopkins University Press, London.

Comment

by
Jérôme Pasquier
Deputy Director General of the DGCID,
French Ministry of Foreign Affairs

I will not comment on all the rich aspects of Mr. Gunning's presentation, in order to better:

- suggest, in all modesty, a path of complementary research,
- highlight the topicality of his words, using the example of recent developments in French aid management since last summer.

1. Proposal for a path of complementary research

Official development assistance is an essential resource for fighting poverty, but we must also realise that other policies led by developed countries sometimes have even more impact.

As Mr. Gunning recalls, the DFID stresses the fact that, in terms of poverty reduction, these other activities can actually cost less than development aid (which in my opinion is focused a little too exclusively on the fight against poverty).

For example, in addition to aid and trade policy, let us look at migratory policy. The impact that migration has on the development of countries of origin has been the object of abundant research, and the issue is clearly important for both its negative (such

as brain drain) and positive effects (such as the transfer of resources by migrant workers). Also important are environmental policy, research and development, attitudes toward foreign investment, etc. To this, we could also add education (with a view to sensitise public opinion to the importance of ODA).

While analysts are attempting to bring aid and development closer together, their analyses are not taking into account the other effects that policies have on developing countries. Are these effects likely to mortgage the aid's own effects?

Authorities responsible for managing aid should assume full responsibility for observable impacts in developing countries. Since 1991, the OECD has been pushing to the forefront the notion of *policy coherence for development*—a theme developed during the nineteen nineties, which contributed to both the Millennium Declaration and the Monterrey Consensus regarding development financing. We must realise, however, that implementation is still in its infancy.

If the literature on aid emphasises governance in beneficiary countries as an effectiveness factor, could we not aim the study in another direction and examine the donor countries' policy coherence under the governance lens?

The developed countries are not very inclined to take these "secondary effects" into account as much as their national interest. This would, however, strengthen governance in the developed countries and contribute to a subsequent stage of state reform on our side of the equation. But are public opinions ready?

Aid donors should therefore argue that aid becomes a political tool amongst others—all of which should be mindful of collective well-being, not merely within national borders but also in poor countries. Research can work to support this perspective by highlighting the effects attributable to our past mistakes and incoherence of our policy.

In France, considerable progress could be made with the current reform, initiated under the CICID (Interministerial International Co-operation and Development Committee) framework of July 20, 2004. It allows for a possible move to coherence by conferring an important place on a delegate minister:

"Under the authority of the foreign minister, the delegate minister in charge of co-operation and development co-ordinates the different actors of the French co-operation programme, monitors the proper realisation of expected results, and regularly informs the President of the Republic and the government of the realisation of quantitative and qualitative objectives."

In my opinion, this represents a first step toward coherence because the delegate minister is clearly in charge of a co-ordination mission—and this could extend to all policies because the CICID brings together many members of government.

2. Aid in France in 2005

France works actively to increase aid resources. As you know, we support the English IFF initiative, which aims at issuing publicly guaranteed bonds. We also work actively at promoting additional resources such as international fiscality. We expect the year 2005 will see us implementing these approaches.

The resulting growth in resources could help form the critical mass necessary to make change credible and initiate an equitable process. We could then avoid the deception that Gunning describes when he shows the limits of conditionality.

It will, of course, take time, but the framework for concentrating aid proposed by the Millennium Development Goals does present a way to bring about these massive effects starting today. However, even with this, the aid agenda is far from completed.

2.1. Do the MDGs make it possible to avoid unfavourable political economy for donors in relation to beneficiary governments?

While the political economy of development aid schematically opposes cynical beneficiary governments to angelic donor governments (whose objectives nevertheless tend towards a higher rate of aid mobilisation), there is a way to avoid this game: via resolute alignments between donors on a small number of themes, which in turn makes credible the idea that visible, decisive results can be obtained. Thus the Millennium Goals can be a part of the answer, because they have become a politically

unavoidable reference in the development and aid discourse, and because all the donors refer to it.

However, this politically mobilising approach is intellectually lacking and simplistic.

It does mobilise—in France the CICID decided on July 20, 2004 to put in place sector strategies in the domains closely linked to the implementation of the Millennium Development Goals. French aid has even drawn conclusions from the growing role of the Millennium Goals by increasing its efforts in primary education, basic health care and the fight against AIDS.

But if we lose sight of the fact that this motivation stemmed from a preoccupation with communication, then the approach via Millennium Goals is lacking—from the standpoints of both geography and development dynamics.

Looking at the diversity of regional situations and needs of the beneficiaries, if we keep a universal approach and comprehensive measuring, the success or failure of the Millennium Development Goals will owe much to the populous countries of South and East Asia. If we prefer a more differentiated vision, the problem of the Millennium Goals appears to be essentially African. What we gain in exactitude, we lose in message universality.

While it has become accepted that the Millennium Development Goals should be a framework for mobilisation around which national representatives (citizens, NGOs and members of parliament) and large aid agencies interact, they are far from being an accurate reflection of what should be done from the general and French point of view.

2.2. The development agenda is wider

The Millennium Development Goals do not exhaust the development agenda, nor can they aspire to exhaust the international Cupertino agenda.

The consensus around the MDGs has been an element of the “re-legitimisation” of official development aid. Keeping the magnitude of the need in mind, this is a positive point. For all that, can we say that the recent increase in ODA flows is a result of the

adoption of MDGs? Not for sure. What proportion of this global financial increase is due to crisis situations?

The difficulty in mobilising classic ODA explains the researching of innovating finance mechanisms. While the Millennium Development Goals bring consensus, innovative or classic methods of aid financing do not.

The financing of needs linked to the MDGs in social sectors has revived the debate over grants/loans in favour of grants, which are more adapted to finance this type of needs. The polarisation over the "social" content of the MDGs is pushing in favour of the abandonment of loans. As for innovative financing, the mere fact that the subject remains on the international agenda is a success in itself, as its political acceptability is still uncertain. Finally, mechanisms that resort to financial engineering, such as guarantees, are subject to bureaucratic obstacles from the DAC, which refuses a posting, except in cases of failure when guarantees are called—which does not prove failure in terms of development. In my opinion, this seems paradoxical.

Besides the vision of a minimum of social protection, which is the fabric of Millennium Development Goals, the growth and development imperative must remain present in operational methods and resource mobilisation. Growth is good for the poor, and thus for the Millennium Development Goals.

The July CICID meeting brings up to date a multi-dimensional vision of the French development approach: "Along with the Millennium Goals, France will pursue its co-operation in the traditional domains of governance, public policy support, promotion of cultural diversity, higher education, and research." While the last UNDP report on human development shows the importance of cultural liberty, this dimension remains insufficiently recognised. Does it not, however, introduce a little complexity and effectiveness into the game between donors and partner countries?

2.3. From MDGs to global public goods

The current work by the Franco-Swedish task force on Global Public Goods (GPG) can lead to a renewal of the international Cupertino agenda beyond the paths evoked by Mr. Gunning.

The task force has invoked the following themes as a potential basis for the provision of global public goods: (1) peace and security, (2) international trade, (3) international financial stability, (4) sustainable management of natural resources, (5) prevention and control of infectious diseases, (6) knowledge and research. From a thematic point of view, this agenda is of course not disconnected from that of the MDGs, i.e. they can constitute prerequisites (peace and security, financial stability) or they can fit in naturally (sustainable management of natural resources). However, the GPG approach is not limited to the MDG agenda and in certain cases it raises new issues: the question of education is not seen as only basic education, but more so as access and sharing of knowledge (thus the role given to higher education and research).

The relevance of this agenda in terms of regulation of globalisation is undeniable. This also the case for development needs. Without a doubt, middle-income countries (and, within these, only a part of the population) are more concerned than low-income countries under stress (LICUS). However, from the financial crisis of 1997 to that of SARS, from debates on intellectual property protection to those of European Union peace-building initiatives, the topicality of the GPG agenda becomes more apparent.

The questions of governance, capacity building, the role of institutions and public policy must gain a more recognised place on the international co-operation and development agenda. It has less to do with developing new conditionalities than recognising the fundamental role public policy has in stimulating economies and encouraging populations to participate in the decision-making process.

This agenda is, to a large extent, that of the new international middle class—the emerging countries. For these countries, the international agenda of official development aid is only partially relevant. The issue is not—or is no longer—the social treatment of poverty reduction. Some will have to face up to the question of social treatment of inequalities, but whether this subject is a matter of official development aid is under debate. It implies first and foremost a response in terms of public policy.

The interests of emerging countries do not coincide necessarily with those of the least developed countries (LDCs), but they have become imperative. This is particularly clear in multilateral trade negotiations and, more and more, in the field of access and sharing of knowledge.

This agenda is partly that of the French co-operation programme, notably that of the Ministry of Foreign Affairs. By demanding the design of an emerging country assistance strategy, the CICID meeting of July 20 has taken note. Co-operation with these countries falls into the scope of a logic of influence, of “laboratories of change”. It does not require massive financing or necessarily imply steps toward harmonisation. I must, however, support a vision of “regulated globalisation”, an idea backed by France, notably with countries that are more attracted by multipolarity that by multilateralism.

Reforming the Formula: A Modest Proposal for Introducing Development Outcomes into IDA Allocation Procedures

by

Ravi Kanbur*

T.H. Lee Professor of World Affairs, International Professor of Applied Economics and Management, and Professor of Economics, Cornell University.

www.people.cornell.edu.pages.sk145

1. Introduction

How should aid donors allocate aid between recipient countries if their objective is to advance development?[1] This question poses both conceptual and operational issues. All donors have rules and procedures underlying the determination of the level and composition of aid transfers to recipients. In many cases there is an explicit formula which, while not determining in a mechanical sense, certainly sets the benchmarks from which the allocation decision begins. One such formula is the IDA allocation formula, though other donors have procedures that are similar in spirit.

* This paper was presented at the AFD-EUDN conference in Paris, 25-27 November 2004. The ideas in this paper have been presented at seminars and panels at Princeton University, IFAD (Rome), IFPRI (Washington, DC), the World Bank (Panel on Lessons of the 1990s), the DPRU/TIPS/Cornell conference on African Development (Cape Town), and at DFID's Conference on Reaching the Very Poorest (London). Parts of this paper draw on my contribution to the DFID conference, "What Change Does Attention to the Poorest Imply?" I am grateful to participants at these meetings for their helpful comments.

1. For overviews of the literature on aid, see Tarp (2000) or Kanbur (2003).

A very simple framework would suggest the importance of two key factors in choosing between potential recipient countries. First, how successful would this aid be in furthering development? Second, how is development in one country to be valued against that in another? The first is an "aid productivity" question. The second is a "valuation of outcomes" question. The second question is relatively easy to answer if the donor's valuation of development in recipient countries is clear. Given the development outcomes of interest to the donor (e.g. a reduction in infant mortality), a natural specification of the valuation is that a unit improvement should be of greater value the lower the starting point. Thus, roughly speaking, for any given degree of aid productivity, aid allocation should vary inversely with the country's level of development (the exact relationship would need a closer specification of the valuation function).

The issue of valuation of development outcomes is not without its complexities.[2] However, it can be argued that, at least to some extent and especially in the wake of the consensus on the Millennium Development Goals (MDGs), the international community has certain notions of what it values as the outcomes of development. Rather, it is the first question that has vexed aid analysts and practitioners alike, because the productivity of aid is not independent of the modalities of aid delivery and the use of that aid. The arc of thinking has traversed a project-oriented phase, during which the outcomes of specific projects supplied the guide to aid allocation, and a policy-oriented phase, during which the policy parameters of the recipient country were seen as a better guide to the productivity of aid. The discussion has often been cast in terms of the much used, and abused, term "conditionality".

At its most general, conditionality is nothing more than the rules and procedures according to which a donor transfers resources to a recipient. To oppose conditionality in general does not make sense. The devil really is in the detail of the rules and procedures according to which aid is allocated and disbursed.[3] These rules and procedures operate at different levels: in the overall resource envelope allocated to a country, in the division of this envelope between different types of assistance (for example projects or programmes), and in the specific conditions that apply to particular projects or programmes.

2. See Kanbur (2004a).

3. For a discussion of conditionality in the context of the history of development assistance, see Kanbur (2003).

This paper addresses the logic used in determining a country's allocation of its overall aid resource envelope. Since total resources are finite, such allocation has to be based, explicitly or implicitly, on a comparison of relevant features of different recipient countries. Perhaps the most prominent method for comparison is the IDA allocation formula, not simply in terms of the total volume of resources that it allocates, but also because it is generally acknowledged that IDA procedures exert a strong influence on those of other donors. The component that is of specific interest in this paper is the method of cross-country comparison, the Country Policy and Institutional Assessment (CPIA) formula. The paper considers the logic of this formula, before proposing a revision of it.[4]

The structure of the paper is as follows. Section 2 outlines the IDA allocation procedure and the role of the CPIA in this procedure. Section 3 discusses the logic behind the use of the CPIA and offers a critique. Section 4 proposes allocations based on development outcomes and debates the major criticisms of this approach. Section 5 concludes by offering a modest revision of the CPIA as the first step to moving towards a development outcomes-based approach.

2. Outline of the IDA formula[5]

At the core of the logic of the IDA allocation process is a balance between "needs" and "performance". Needs are measured in a straightforward manner by national income per capita, GNIPC. Performance is measured by a performance rating, PR, which is the focus of this paper. The allocation per capita for a country is a function of GNIPC and PR. In fact, the specific relationship is (World Bank 2003a):

$$\text{Allocation per capita} = f\left(PR^{2.0}, GNIPC^{-0.125}\right)$$

4. There are, of course, many aspects of the development assistance process that are important but are not covered in this paper. For example, the sometimes perverse incentives in aid agencies to move money rather than focus on the best use of that money, or the interplay between foreign policy objectives and development objectives in the *realpolitik* of development assistance allocations. Also, my specific focus is on IDA, so I will not be discussing formulae used by other agencies such as the European Union, DFID or USAID.

5. The procedures and the formula are summarised in World Bank (2003a) World Bank (2003b) and updated in World Bank (2004a).

Thus, the performance rating is raised to the square power and per capita income is raised to a negative power, minus 0.125, and these two are then combined to decide the allocation. The function f is chosen to reflect the fact that individual country allocations have to add up to the total resources available. One feature of note is that the performance rating has a much higher weight than the needs measure. This however, is not the chief concern of this paper. Rather, the focus is on how the PR index is constructed and the logic underlying this construction.

Before turning to the PR index, some further clarifications are in order on how the above formula is used. The allocation per capita derived above is not a hard and fixed amount, but rather a "norm". The detailed determination of the allocation, and of the composition of this allocation between different types of assistance, takes place in the Country Assistance Strategy (CAS). To quote World Bank (2003a):

"The allocation norm establishes the financial resources available for each IDA country for the following three fiscal years. The allocation sets the resource envelope that each country could expect to receive if its performance stays the same and assuming a pipeline of quality projects—but is not an entitlement. In the case of a new CAS, the allocation norm will set the base-case financing scenario…The CAS financing scenarios may be adjusted to reflect special country circumstances, which will be spelled out in the CAS."

Moreover, there are a number of exceptions to the norm derived above. The report continues:

"In addition to their performance-based allocations, all countries are allotted a basic allocation of SDR 3 million (about US $4 million). In terms of per capita allocations, this benefits in particular the small states. There are some important considerations that merit exceptions to the allocation norms. First, "blend" countries with access, or potential access, to IBRD receive less than their norm allocation due to their broader financing options. Second, post-conflict countries can, when appropriate, be provided with additional resources in support of their recovery and in recognition of a period of exceptional need. And third, additional allocations may be provided in the aftermath of major natural disasters."

However, despite these caveats, the allocation norm, and the performance rating that underlies it, is a central feature of the whole process.

Table 1.
2003 CPIA Categories

A. Economic Management
1. Management of Inflation and Macroeconomic Imbalances
2. Fiscal Policy
3. Management of Public Debt (External and Domestic)
4. Management and Sustainability of the Development Programme
B. Structural Policies
5. Trade Policy and Foreign Exchange Regime
6. Financial Stability
7. Financial Sector Depth, Efficiency and Resource Mobilisation
8. Competitive Environment for the Private Sector
9. Goods and Factor Markets
10. Policies and Institutions for Environmental Sustainability
C. Policies for Social Inclusion/Equity
11. Gender
12. Equity of Public Resource Use
13. Building Human Resources
14. Social Protection and Labour
15. Monitoring and Analysis of Poverty Outcomes and Impacts
D. Public Sector Management and Institutions
16. Property Rights and Rule-based Governance
17. Quality of Budgetary and Financial Management
18. Efficiency of Revenue Mobilisation
19. Quality of Public Administration
20. Transparency, Accountability and Corruption in the Public Sector

Source: World Bank (2003b)

How is the PR index derived? At its heart is the Country Policy and Institutional Assessment (CPIA). The procedure for 2003 was as follows (the 2004 procedure was slightly modified as noted below). Essentially, this is an assessment of a country on each of the 20 items divided into four categories, as shown in Table 1. Each of these items is then scored by Bank staff on a scale from 1 (low) to 6 (high). The broad interpretations of these scores are given in Table 2. The specific guidelines are elaborated in the 2003 CPIA questionnaire:

"Countries should be rated on their current status in relation to these guidelines and to the benchmark countries in each region, for which the agreed ratings have been provided to the staff. Please assess the countries on the basis of their currently observable policies, and *not* on the amount of improvement since last year, *nor* on intentions for future change, unless the latter are virtually in place…As described in these guidelines, a "5" rating corresponds to a status that is good today. If this level has been sustained for three or more years, a "6" is warranted, signifying a proven commitment

Table 2.
2003 CPIA Ratings Scale

1 (low) through **6** (high)

1	Unsatisfactory for an extended period
2	Unsatisfactory
3	Moderately Unsatisfactory
4	Moderately Satisfactory
5	Good
6	Good for an extended period

Intermediate scores of 2.5, 3.5 and 4.5 may also be given.
Scores of 1.5 and 5.5 may not be given.

Source: World Bank 2003(b).

Table 3.
2003 IDA Country Performance Ratings

First Quintile	Armenia, Benin, Bhutan, Cape Verde, Ghana, Grenada, India, Madagascar, Mauritania, Samoa, Sri Lanka, St. Lucia, St. Vincent and the Grenadines, Tanzania, Uganda
Second Quintile	Azerbaijan, Bosnia and Herzegovina, Burkina Faso, Dominica, Guyana, Honduras, Malawi, Mali, Mozambique, Nepal, Nicaragua, Pakistan, Rwanda, Senegal, Serbia and Montenegro, Republic of Yemen
Third Quintile	Albania, Bangladesh, Bolivia, Cameroon, Ethiopia, The Gambia, Indonesia, Kenya, Kyrgyz Republic, Lesotho, Maldives, Moldova, Mongolia, Vietnam, Zambia
Fourth Quintile	Burundi, Chad, Democratic Republic of Congo, Republic of Congo, Côte d'Ivoire, Djibouti, Eritrea, Georgia, Guinea, Kiribati, Lao PDR, Niger, Sao Tome and Principe, Sierra Leone, Tajikistan
Fifth Quintile	Angola, Cambodia, Central African Republic, Comoros, Guinea-Bissau, Haiti, Nigeria, Papua New Guinea, Solomon Islands, Sudan, Togo, Tonga, Uzbekistan, Vanuatu, Zimbabwe

Countries not rated in 2003 exercise: Afghanistan, Liberia, Myanmar, Somalia, and Timor-Leste.
Note: Number of countries per quintile varies due to equal scores at cut-off levels.
Source: World Bank (2003c).

to and support for the policy. Similarly, a "2" rating represents a thoroughly unsatisfactory situation today. A "1" rating signifies that this low level has persisted for three or more years, and therefore that the resulting problems are likely to be more entrenched and intractable." (World Bank, 2003b)

Finally, a simple unweighted average of these scores is taken to give the CPIA index. Individual country scores are not released to the public, only country quintiles are made available (this is slated to change in 2005). The results for 2003 are given in Table 3.

Before turning to the specific categories and their scoring criteria, it is worth specifying how the CPIA feeds into the PR. First the CPIA is combined with the Bank's Annual Review of Portfolio Performance (ARPP), the weights being 80% for CIPA and 20% for ARPP. Then this weighted average is multiplied by a "governance factor". The governance factor is built up as follows. First, an unweighted average is taken of the scores for six governance-related criteria in the CPIA, numbers 4 and 16-20 (see Table 1), and of a seventh score, on the "procurement practices" criterion from the ARPP assessment process (since it is not the focus of this paper, the ARPP process is not

Table 4.
2003 CPIA: Score Guidelines for Fiscal Policy

2	Fiscal balance will likely lead (or is already leading) to inflationary financing, crowding out of private sector investment, an unsustainable current account deficit or an unsustainable level of public debt; or fiscal policy is not making a serious attempt at provision of public services and infrastructure essential to growth.
3	Sporadic efforts at macroeconomic stabilisation through fiscal policy but not maintained consistently or implemented through temporary measures like ludicrously low real public sector wages or cuts in projects or services with high long-run returns; or attempts at public services and infrastructure provision are sporadic and concentrated in not very cost-effective uses of funds.
4	Consistent maintenance of macroeconomic stability and fiscal sustainability through appropriate levels of the fiscal balance and lasting adjustment measures with only occasional lapses. Public service provision is good in some sectors but still inadequate in others.
5	Fiscal policies are consistent with adequate provision of high quality public services and infrastructure for economic growth and generate a fiscal balance that can be financed (including with aid flows where applicable) in a non-inflationary way and is consistent with adequate credit for the private sector and a sustainable path of public debt.

Source: World Bank 2003(b).

Table 5.
2003 CPIA: Score Guidelines for Trade Policy and Foreign Exchange Regime

2	Average tariff (weighted by global trade flows) is high (over 30%). High and erratic import and/or export barriers, including quantitative restrictions and/or state trading monopolies. Export taxes or quantitative restrictions frequently used. Customs or political authorities make discriminatory or ad hoc exemptions. Valuation procedures arbitrary and artificial exchange rates result in substantial over or under valuation of goods for customs purposes. Clearance of goods requires many approvals, arbitrary fines, frequent bribes to customs officials and involves long delays. Foreign exchange rationed or an administered foreign exchange regime with multiple exchange rates.
3	Average tariff 20-30%. Coverage of quantitative restrictions reduced to 15% or lower. Export restrictions mostly phased out. Duty exemptions frequently used to offset the adverse impacts of import barriers on inputs used in production of exports or for approved investment projects. Foreign exchange convertible for most current account purposes. Customs clearance involves high inspection ratios for imports, and interaction between officials and importers encourages corruption.
4	Average tariff 10-20%. Quantitative restrictions cover only a very small percentage of imports (under 5%). Virtually no export restrictions. Duty exemptions provided only in accordance with well-defined rules. Foreign exchange convertible for virtually all current account purposes. Customs procedures are well defined, quick, efficient and impersonal, and staff professional, although some "tea money" payments to expedite clearance may still be present.
5	Average tariff (weighted by global trade flows) is low (10% or less), with low dispersion and insignificant or no quantitative restrictions or export taxes. Trading monopolies absent or unimportant. Indirect taxes (e.g. sales, excise, surcharges) do not discriminate against imports or exports. Efficient and rule-bound customs administration. IMF Article 8 status. Minimal or no foreign exchange restrictions on long-term investment capital inflows.

Source: World Bank 2003(b).

discussed in any further detail). This average score is then divided by 3.5 (the mid-point of the 1-6 scoring range), and this ratio is raised to the power of 1.5. This procedure effectively ends up giving significantly greater weight overall to the governance criteria in the CPIA.[6]

The components of the CPIA are thus central building blocks in the whole process. There are specific guidelines for the scoring of each of the 20 items that make up the CPIA. Tables 4, 5, 6 and 7 lay out these guidelines for one component from each of the four major categories in the CPIA: Fiscal Policy under Economic Management, Trade Policy and Foreign Exchange Regime under Structural Policies, Equity of Public

6. Note that this is the procedure for 2003. For 2004, a revised procedure was adopted, as set out in World Bank, 2004a.

Table 6.
2003 CPIA Score Guidelines for Equity of Public Resource Use

2	Most public expenditures for economic and social services do not benefit the poor more than the better off. The government has not identified individuals, groups or localities that are poor, vulnerable, or have unequal access to services and opportunities, does not have appropriate programmes, and has no plans. Spending on economic and social services targeted to the poor is inadequate. The overall incidence of revenues is regressive.
3	Only some public expenditures for economic and social services benefit the poor more than the better off. The government has not identified individuals, groups or localities that are poor, vulnerable, or have unequal access to services and opportunities, does not have appropriate programmes, and has taken only small, if any, steps to correct this. Spending on some key economic services targeted to the poor is inadequate. The overall incidence of revenues is regressive and only small steps, if any, are being taken to correct this.
4	Key public expenditures for economic and social services benefit the poor more than the better off, but some egregious regressive expenditures remain. The government has identified individuals, groups or localities that are poor, vulnerable, or have unequal access to services and opportunities, and is taking significant steps to introduce appropriate programmes. With few exceptions, spending on economic services targeted to the poor is broadly adequate. The overall incidence of revenues is progressive, but some egregious regressive revenue sources remain.
5	Key public expenditures for economic and social services are well targeted to benefit the poor. There are few, if any, egregious regressive expenditures. The government has identified individuals, groups or localities that are poor, vulnerable, or have unequal access to services and opportunities, and has designed appropriate programmes. Spending on economic services targeted to the poor is broadly adequate. The overall incidence of revenues is progressive, and there are few, if any, egregious regressive revenue sources.

Source: World Bank 2003(b).

Resource Use under Policies for Social Inclusion/Equity, and Transparency, Accountability and Corruption in the Public Sector under Public Sector Management and Institutions. Note that guidelines are specified only for scores of 2 (unsatisfactory), 3 (moderately unsatisfactory), 4 (moderately satisfactory), 5 (good); a score of 1 is simply "unsatisfactory for an extended period" and a score of 6 is "good for an extended period".

Finally, we note that in 2004, certain changes to the CPIA process were accepted by World Bank management (see World Bank, 2004a). Among these is the intention to disclose CPIA scores from 2005 and to establish an independent expert standing committee to review the CPIA methodology every three years. These movements are to be highly welcomed. In addition, the governance factor calculation was changed,

Table 7.
2003 CPIA Score Guidelines for Transparency, Accountability and Corruption in the Public Sector

2	There are no effective audit or other checks and balances on executive power. Boundaries between the public and private sector are ill-defined, and conflicts of interest abound. Responsibilities are not clearly defined across levels of government and the reasons for and costs of decisions by public officials and the judiciary are not made clear or are not based on legal rules or procedures. Laws and policies are biased towards narrow private interests, implementation of laws and policies is distorted by corruption and resources budgeted for public services are diverted to private gain. The media are not independent of government or powerful business interests. Public officials are not sanctioned for failures in service delivery or for receiving bribes. The general public has little voice or participation in public activities.
3	Elected and other public officials often have private interests that conflict with their professional duties. Decision making is generally not transparent. External accountability mechanisms such as inspector-general, ombudsman or independent audit may exist, but have inadequate resources or authority. Restrictions on the media limit its potential for information-gathering and scrutiny, and civil society is weak.
4	External accountability mechanisms limit somewhat the degree to which special interests can divert resources or influence policy making through illicit and non-transparent means. Media publicity is an effective deterrent against unethical behaviour. Risks and opportunities for corruption within the executive are reduced through adequate monitoring and reporting lines. Conflict of interest and ethics rules exist and the prospect of sanctions has some effect on the extent to which public officials shape policies to further their own private interests. Administrative corruption is low.
5	Responsibilities are clearly defined across levels of government. Accountability for decisions is ensured through a strong public service ethic reinforced by audits, inspections, and adverse publicity for performance failures. The judiciary is impartial and independent of other branches of government. The reasons for decisions, and their results and costs, are clear and communicated to the general public. Citizens can obtain government documents at nominal cost. Conflict of interest and ethics rules for public servants are observed and enforced. Top government officials are required to disclose income and assets, and are not immune from prosecution under the law for malfeasance. Authorities monitor the prevalence of corruption and implement sanctions transparently.

Source: World Bank 2003(b).

and the number of CPIA categories was reduced to 16, as given in Table 8. However, albeit with new categories, and a new procedure for calculating the governance factor, the essence of the CPIA method and the IDA allocation formula are left unchanged.

This completes the outline description of the IDA formula, and its centrepiece, the CPIA scores. What is the logic underlying this method of aid allocation? It is to this question that we now turn.

Table 8.
Revised 2004 CPIA Categories

A. Economic Management
1. Macroeconomic Management
2. Fiscal Policy
3. Debt Policy
B. Structural Policies
4. Trade
5. Financial Sector
6. Business Regulatory Environment
C. Policies for Social Inclusion/Equity
7. Gender Equality
8. Equity of Public Resource Use
9. Building Human Resources
10. Social Protection and Labour
11. Policies and Institutions for Environmental Sustainability
D. Public Sector Management and Institutions
12. Property Rights and Rule-Based Governance
13. Quality of Budgetary and Financial Management
14. Efficiency of Revenue Mobilisation
15. Quality of Public Administration
16. Transparency, Accountability, and Corruption in the Public Sector

Source: World Bank (2004a).

3. The logic of the formula, and a critique

There are many specific and operational criticisms of the IDA allocation process. The CPIA is carried out behind closed doors by Bank staff, with little or no scrutiny from outside independent observers (scheduled to change in 2005). The ARPP remains an under-scrutinised assessment procedure, linked as it is to internal Bank procedures. The way the "governance factor" enters the formula is convoluted at best. And, the origin of the different weights and exponents used in various parts of the formula is not clear. Why, for example, is PR raised to the power 2, while the governance score ratio is raised to a power of 1.5 to give the governance factor? Why exactly is GNIPC raised to the power of minus 0.125? The main concern in this paper is not with these specifics; any formula will have to make such operational specifications and defend them the best it can. Rather, our concern is with the fundamental logic of the process.

As noted in the introduction, any logic for allocating development assistance resources to a poor country must include two elements: how much the assistance can be translated into improvements in outcomes important to the donor ("aid productivity" or "performance"), and how much the donor values these improvements in outcomes ("need"). Thus, if D is a measure of the final development outcomes and W(D) is the donor's valuation of it, then the impact of aid A can be written mathematically as:

$$dW/dA = [\delta W/\delta D] \times [\delta D/\delta A] = \text{Need For Aid} \times \text{Productivity of Aid} \quad \textbf{(1)}$$

The first term on the right-hand side values development outcomes as seen by the donor, while the second term measures the impact of a unit of aid on development outcomes, *i.e.*, its productivity. If the value the donor places on the outcomes declines as the outcome improves, then the need dimension can be captured by an inverse function of the level of the desired variable. In the IDA allocation formula, this is done simply by taking the per capita national income of a country and raising it to the power of minus 0.125. Wealthier countries will get lower allocations through this component of the formula. Thus the IDA formula essentially captures need through the income criterion, and does not go directly to indicators such as infant mortality, maternal mortality, girls' education and other components of the Millennium Development Goals (MDGs), through which the international community has presumably expressed its objectives of the development process. However, we will set this aspect of the IDA formula to one side, since our main focus is on the way that performance is measured.

Conceptually, if we hold the needs part of the formula constant, then more aid should flow where its impact on objectives is greatest. If we could identify environments in which aid productivity is highest, *i.e.* where improvements in final development outcomes of interest, per unit of aid flow, would be greatest, then more aid should be allocated to those environments. Presumably the performance rating part of the IDA formula, and specifically the CPIA component of it, attempts to identify high aid productivity environments. The logic must be that a higher score on any of the 20 components of the CPIA enhances the productivity of aid flow and therefore argues for more aid. These scores are then aggregated with equal weights cross the 20 categories. There are two possible logics behind this last step. Either these categories are equally valuable to aid productivity, or it could be that we have no information on the relative contribution

of each category to overall aid productivity so, on the principle of insufficient justification, each category should be given equal weight.

But perhaps the most striking, yet least noticed feature of the PR formula, and especially the CPIA, is that *it is the same for every country*. The 20 categories do not change, the scoring guidelines in each category are the same, and the weighting scheme (equal weighting) is the same for every country. What is the logic behind this uniformity in country treatment?

One way to access the logic is to consider the literature on "cross-country growth regressions", not least because this literature has had a tremendous influence in thinking on development strategies and aid strategies. In this literature, economic growth in a country is seen as a function of a number of determining variables. If growth rate is G then growth in country i is given by:

$$G_i = \alpha + \beta Y_i + \theta X_i + \gamma A_i + \eta X_i A_i + \epsilon_i \qquad \textbf{(2)}$$

where Y is a vector of structural variables that the government cannot control (such as a country's geography and climate), X is vector of policy variables (such as fiscal deficit, tariffs, percentage of government expenditure devoted to primary education, or independence of judiciary) and ϵ is a classical stochastic error term. The coefficients in α, β, θ, and η translate the impact of their respect variables to growth. Thus, according to the world view implicit in Equation 2, a country's growth depends upon structural features that the government cannot control, policy variables that the government can control, aid flows, and an interaction term between aid flows and these same policy variables. This is, sometimes quite literally, the family of regressions that have been run over and again in the literature, including the well-known contribution of Dollar and Burnside (2000), and subsequent papers by other authors.[7] While a relationship like Equation 2 is most often estimated with growth as the dependent variable, there is no reason why it could not be used to explain other dependent variables such as infant mortality or life expectancy. To do this, we simply replace G_i with Di.

7. See, for example, Hansen and Tarp (2000), Dalgaard and Hansen (2001), Guillaumont and Chauvet (2001), Easterly, Levine and Roodman (2003). See also the survey in Kanbur (2003).

While there are a number of data and econometric problems associated with estimations such as Equation 2, they, again, are not the main focus and I want to set those aside for now. Of note, however, is that Equation 2 sees no role for the aid flows themselves to influence policy. In other words, it leaves no role for conditionality in changing government policy. This is surely right because, if the experience of two decades has taught us anything, it is that the development assistance tail cannot wag the domestic political economy dog. Rather, we should take the policies emerging out of the domestic political economy, and accept them as given in aid allocation decisions.[8] If the relationship underlying Equation 2 does indeed hold in a cross section of countries, then the implications for aid allocation are clear, since in mathematical expectation,

$$\delta G_i/\delta A_i = \gamma + \eta X_i \qquad \textbf{(3)}$$

The productivity of aid is then given by the right-hand side of Equation 3, where the values of the policy variables in country i, the elements of the vector X_i, weighted by the elements of the vector η, (which are estimated from the regression run on cross-country data) are all stated. The logic of the IDA formula is now clear. Equation 2, and its derivative, Equation 3, lead to the scoring function given in the right-hand side of Equation 3. This informs us which policies should be counted (the elements of X_i) and how they should be weighted (the elements of η).

Having laid bare the logic, let us consider it further. The basic point is that the scoring rule in Equation 3 is only as valid as the underlying model in Equation 2 and its econometric estimation. First, it is uncertain whether a regression has previously been run with the 20 policy categories in the IDA formula using the scores for policies as exemplified in Tables 4-7 to generate the elements of X_i. Even if it has, it is almost certain that the results would not show the elements of η all being equal, thereby giving the equal weighting rule. Rather, the categories in the IDA formula reflect an accretion of factors thought to be important to the development process under different arguments made in different contexts. The equal weighting of the different factors then really reflects the "principle of insufficient reason", rather than a reasoned logic leading to particular combination of key policy factors that impact on the productivity of aid.

8. In fact, in Burnside and Dollar (2000) a jointly estimated equation testing for the impact of aid on policies finds no such relationship. On conditionality, there is of course a huge literature. For example, see Guillaumont and Guillaumont (1995), Kanbur (2000, 2003), and Adam *et al.* (2003).

But perhaps most important is the fact that a common scoring rule for all countries, which exists in Equation 3, depends upon a common development-outcomes model for all countries. Put differently, Equation 2 effectively assumes that *all* factors explaining outcome variations across countries have been successfully accounted for in the variable included in Equation 1. Also the effect of an explanatory variable on development outcomes is *identical* across countries. Any variation across countries over and above that accounted for by the explanatory variables is purely random, not amenable to further parsing.

Over the past decade, dissatisfaction has been growing with the estimation of a cross-country "average relationship" leading to "best practice" policy guidelines common to all countries. The view that variations around the estimates of average relationships in Equation 2 are not simply random variations, but reflect country specific factors captured neither in the model nor in the data, was powerfully put in a recent report from the World Bank itself, *Economic Growth in the 1990s: Learning from a Decade of Reform*:

"The Study concludes that valid general principles do not imply generic "best practice" policy or institutional solutions…Regarding macroeconomic policies for example, the findings emphasise the importance of institutions underlying macroeconomic stabilisation, the risks associated with external financial liberalisation, the disruptions associated with episodes of exchange rate appreciation, and the sometimes excessive focus on minimising inflation in the short term…Regarding trade, the analysis highlights the fact that countries that have successfully integrated into the world economy have followed different approaches and also adopted a range of complementary policies, making it difficult to pin down the exact relationship between trade integration and growth…Perhaps the lesson of the lessons of the 1990s is that we need to get away from formulae and realise that economic policies and institutional reforms need to address whatever is the binding constraint on growth, at the right time, in the right manner, in the right sequence, instead of addressing any constraint at any time…"[9]

No doubt this view will be debated heatedly both within the World Bank and in the development community at large. But let me record here that I support the "end of

9. World Bank, 2004b, pp. vi-vii.

certainty" heralded by this report. In this context, then, what I want to highlight again is that the CPIA does not contain any final outcome variables like poverty, extreme poverty, girls' enrolment, maternal mortality rates, infant mortality rates, etc. What it has instead is a series of intermediate variables such as trade policy, regulatory policy, property rights, corruption, etc., which we hope will eventually influence the outcomes in which we are truly interested. In effect, it has an implicit model of the development process that states that if the scores on the categories in the CPIA improve, then development outcomes will improve, or rather, the productivity of aid will improve. Over the years, these categories have broadened and increased to 20, and then most recently decreased to 16, but the basic logic that we have the right and complete model, captured in the CPIA, has not changed. And this is then the model that is implicitly assumed to be valid for every country to which the aid allocation formula is applied.

My contention is that the evidence for universally imposing this implicit common model of the development process supposedly leading to improvement in final outcomes, is weak: both for growth and for development outcomes.[10] Despite the support of the World Bank report (2004b), this may be interpreted as a controversial view, so let me be clear about what I am saying. I am not necessarily questioning that the model implicit in the CPIA is a good representation of the average across countries, although in other contexts I would question this as well. Rather, what I am saying is that the country variations around this average, deriving from myriad country specificities that cannot be captured by outside data and outside observers, are large and complex. How else can we explain the fact that Bangladesh, a country that far outstrips others on improvement in social indicators in the last decade, is nevertheless at the top or near the top of Transparency International's corruption index? How else can we explain that once the fiscal deficit is in a range of, say, 2% of GDP, further reductions do not necessarily contribute to increased investment and growth? How else can we explain that two countries can spend about the same amount on primary education, yet in one country this will result in higher enrolment rates and test scores? These variations are not random, but are attributable to specific local factors that are not captured and perhaps cannot be fully captured in our models. The problem lies not in estimating an average relationship given the data that we have; the problem lies in using this average relationship to make country specific judgments.

10. In addition to World Bank (2004b), see the many references in Kanbur (2004a).

Lest I am misunderstood, let me clarify further. There are certainly extreme situations, such as hyperinflation, a double-digit fiscal deficit ratio, a trade system rife with mutually inconsistent quantity controls, a production sector dominated by highly inefficient state enterprises, extremely low spending on education and health, etc., in which general prescriptions are indeed valid, although even here there may be pace and sequencing issues affecting the general direction. But in "normal" cases of drawing lessons from the average, relationships may well obscure the local specificities that determine the success of policies and interventions.

So, in the face of these critiques of the underlying logic of the CPIA's role in the allocation of IDA across countries, what is one to do? It is to this question that the next section now turns.

4. Outcomes-based aid allocation: criticisms and responses

If the average cross-country relationships that underlie the CPIA procedure, and thus the IDA allocation formula, cannot truly capture the cross-country variations in the productivity of aid, what is the alternative? One is, of course, to try even harder to capture country-specific variations in estimating Equation 2, with better data and more elaborate methods. This is a fruitful line of enquiry, but it is not my focus. Such further detailed investigation may never incorporate all country-specific variations. In any event, while that exercise is ongoing, the specific question about what to do with the IDA formula remains.

The solution I propose, initially in an extreme and pure form, follows directly from Equation 1, and is as follows. The needs side of the story can be measured directly by the levels of the targeted outcomes, but the performance side can be measured by the rate of improvement of these variables over a given period of time up to the point of assessment, suitably normalised by the total aid flow over this period. A country that has a very low level of girls' enrolment in primary schools should get more aid on grounds of need. But a country showing rapid improvement of girls' enrolment from this low level, relative to the aid it is receiving, should get even more. A country that is showing relatively slow rates of improvement should get relatively less.

In a purely conceptual sense, I am arguing for an aid allocation formula that depends solely on outcomes for the very poor: with needs measured by initial levels and performance measured by change.[11] There are of course, many criticisms, conceptual and operational, of this position. I want to review a number of these, and present some responses.

Which outcome variables will be chosen?

This is an important question, and focuses the donor's attention on the true underlying objectives. For all its problems, the MDG process has begun to forge an international consensus on these outcome variables. But, the lack of complete consensus on outcome variables seems to be an odd justification for adopting intermediate variables that are relevant presumably only because they will favourably affect our targeted outcomes. If we do not know which outcomes we seek, how can we know the intermediate variables that give rise to these outcomes?

How will the different outcome variables be combined?

Another good question that requires further discussion and consensus. But note that the 20 (or 16 from 2004 onwards) CPIA scores are also aggregated to form one index, with little justification for the method of aggregation (a simple unweighted sum).

What about the "start-up" problem? A country coming out of conflict, for example, will be penalised because it cannot show a track record of performance on outcome variables.

This is a generic issue irrespective of how we judge performance. As noted in Section 2, the current IDA allocation process does indeed make special provisions for post-conflict countries. There is no reason why these principles cannot continue to be used in an outcome-based approach. For example, other criteria could be used for a set period in these cases.

What about shocks and random events that can lead to poor performance (or good performance) in a year, independent of the government's actions?

11. The idea of using actual outcomes in guiding aid allocations is not at all new, of course. For an argument that leads to this proposition from the earlier literature on the design of conditionality in adjustment lending, see Collier *et. al.* (1997).

Such variations can be handled in principle by taking time averages of the outcome variables and the averaging can be statistically as sophisticated as warranted. It should be noted that shocks are often factored into the subjective judgments that go into the assessments of the sub-categories of the CPIA (for example, if tariff reductions did not go as far as might have been hoped because of a revenue crisis stemming from a decline in commodity prices). Moreover, allowances are made for natural disasters and the like in the current CPIA and similar procedures can surely be maintained.

With this "backward-looking" assessment of performance, can an incoming government not live off the achievements of its predecessor? Should we not be holding each government accountable for its actions, and indeed trying to influence these?

These questions articulate the twin conceits that we from the outside know the specificities of what needs to be done, so we can then judge actions, and that the tail of outside development assistance can wag the dog of domestic political economy. It should be clear from the discussion in the previous section that I am sceptical of both of these propositions.

W*hat about the reverse problem, of a government that comes to power after years of a policy of neglect of the poorest (rather than conflict and collapse), and now wants to address this issue seriously? There is no track record of performance on outcomes, but the government wants to take actions that it claims will lead to improved outcomes. Is this not another "start-up" problem that deserves external support?*

I am sympathetic to this argument (note that it holds just as well with the current CPIA methodology), but at the same time I am wary of repeated "new dawn" claims. But surely we can devise mechanisms to handle these cases by ring fencing limited funds for start-up and then letting performance measurement on outcomes take over, just as we seem to do for the "new dawn" cases under the current methodology.

What about data quality and availability for outcome variables? Is this not notoriously bad? How can we possibly run an aid allocation system based on such incomplete data?

These are all valid points. But, first, data for the current intermediate variables (e.g. property rights) are not problem-free. Second, most importantly, we should invest in

monitoring the variables in which we are truly interested. It is odd that we invest in collecting information for intermediate variables, but not on the outcome variables of ultimate interest. But, of course, one of the reasons for doing so is because the intermediate variables enter the aid allocation formulae! This cycle has to be broken somewhere. Indeed, assistance for monitoring development outcomes for the poorest can be (almost) unconditional in any setting. Interestingly, criterion 15 of the 2003 CPIA is "Monitoring and Analysis of Poverty Outcomes and Impacts". IDA is currently emphasising measurement of development outcomes. Although this exercise is currently seen as external to the CPIA and IDA aid allocation processes (indeed, monitoring as a separate criterion is not one of the 16 categories in the new 2004 CPIA. See Table 8), the emphasis on measurement and monitoring can only help if the objective is to make aid allocations more outcome-based.

If we make aid allocation dependent on performance measured by outcome variables, will there not be an incentive for the government to manipulate this information?

But this is true of any measure on which aid allocation will be based. Measures of Domestic Credit Creation (a classic intermediate variable), for example, are not free of the possibility of manipulation. This merely strengthens the argument for strong, locally based mechanisms of monitoring and evaluation.

Will the exclusive focus on the outcomes detract from an understanding of the development process, for which intermediate variables are necessary?

Nothing I have said prevents us from analysing the development process, or developing models and learning lessons from country experience. And, indeed, we can convey those lessons from one country to the next. Focusing on outcomes, however, prevents the easy temptation of saying that because X worked in country A, we will condition aid on X being done in country B. In fact, it may be that country B tries Y and gets the same or better result. If two countries deliver the same improved outcomes using very different intermediate methods, this is of inherent interest to us as analysts. As aid allocators, however, there is a strong argument for submitting the two countries to the same treatment.

Where, then, do these criticisms and responses leave us? The next section proposes a resolution.

5. Conclusion: a modest proposal

Thus, many of the criticisms of the pure proposal—to make the performance part of IDA's aid allocation formula based solely on improvements in outcomes per unit of aid flow—apply equally to the current IDA process. Though there are reasonable responses to all of the criticisms, they remain nevertheless powerful. These would suggest a carefully modulated move in the direction of outcome-based allocation, learning and improving as we go along.

My proposal is therefore as follows:

While leaving the current IDA allocation methodology essentially intact, IDA should introduce one new category of scoring in the CPIA. This category should evaluate the evolution of an actual development outcome variable up to the present. The choice of variable is open. It will depend on international consensus and on data availability, but surely the elements of the MDGs are likely candidates. Once an outcome indicator is chosen, its evolution should be described using appropriate time-series averaging techniques. The rate of change of the averaged time series, per unit of aid flow, should then be scored in the same way as the other CPIA categories, from 1 to 6. Guidelines should be developed for this scoring, just as guidelines were developed for the current CPIA categories. The equal weighting procedure would still be maintained to give the CPIA score. All other aspects of the current IDA allocation procedure would be left unchanged by this proposal.

I hope that this proposal will be considered practical and achievable. It responds to basic concerns about the logic of the CPIA and the IDA allocation formula, and yet it does not radically depart from the current procedure. But it will require a fair amount of technical work to lay the background for choice of indicator, for the averaging technique and for the guidelines on scoring. This work can be done, but if we hope to have it ready for the next CPIA in 2005, we will have to start soon. After three years, the experience can be assessed, and the possibility of introducing new outcome variables (or indeed dropping them altogether) can be evaluated. The fact that there is to be a standing expert committee on the CPIA (see World Bank 2004b) means that this work can be guided by that committee, which can also facilitate technical discussions and consensus building. I commend this modest proposal to the World Bank.

REFERENCES

Adam, C., G. Chambas, P. Guillaumont, S. Guillaumont-Jeanneney and J. W. Gunning (2003), "Performance-Based Conditionality: A European Perspective", Processed.

Burnside, C. and Dollar, D. (2000), "Aid, Policies, and Growth", *American Economic Review*, September, 4(90): pp. 847–68.

Collier, P., P. Guillaumont, S. Guillaumont and J. W. Gunning (1997), "Redesigning Conditionality", *World Development,* 9(25): pp. 1399-1407.

Dalgaard, C.-J. and Hansen, H. (2001), "On Aid, Growth and Good Policies" *Journal of Development Studies,* (6)37: pp. 17–41.

Easterly, W., R. Levine and D. Roodman (2003), "New Data, New Doubts: A Comment on Burnside and Dollar's 'Aid, Policies, and Growth'", *American Economic Review.*

Guillaumont, P. and Chauvet, L. (2001), "Aid and Performance: A Reassessment" *Journal of Development Studies*, 6(37): pp. 66–92.

Guillaumont, P., and Guillaumont Jeanneney S. (1995), *Propositions pour un nouveau type de conditionalities',* European Commission, DG VIII.

Hansen, H. and Tarp, F. (2000), "Aid Effectiveness Disputed" *Journal of International Developmen,t* 3(12): pp. 375–98.

Hansen, H. and Tarp, F. (2001), "Aid and Growth Regressions" *Journal of Development Economics,* (2) 64: pp. 547–70.

Kanbur, Ravi (2000), "Aid, conditionality and debt in Africa" in F. Tarp (ed.) (2000) *Foreign Aid and Development,* London and New York: Routledge.

Kanbur, R. (2004a), "Growth, Inequality and Poverty: Some Hard Questions", http://www.arts.cornell.edu/poverty/kanbur/GrolneqPov.pdf Accessed 23 October 2004.

Kanbur, R. (2004b), "What Change Does Attention to the Poorest Imply?", Paper presented to the DFID conference on *Reaching the Very Poorest*, London, 12 May 2004.

Kanbur, R. (2003), "The Economics of International Aid", http://www.arts.cornell.edu/poverty/kanbur/HandbookAid.pdf Accessed 23 October 2004.

Tarp, F. (ed.) (2000), *Foreign Aid and Development,* London and New York: Routledge.

World Bank (2003a), "Allocating IDA Funds Based on Performance: Fourth Annual Report on IDA's Country Assessment and Allocation Process", http://siteresources.worldbank.org/IDA/Resources/PBAAR4.pdf Accessed 23 October 2004.

World Bank (2003b), "Country Policy and Institutional Assessment 2003: Assessment Questionnaire", http://siteresources.worldbank.org/IDA/Resources/CPIA2003.pdf Accessed 23 October 2004.

World Bank (2003c), "2003 IDA Country Performance Ratings" http://siteresources.worldbank.org/IDA/Resources/QuintilesCPIA2003.pdf Accessed 23 October 2004.

World Bank (2004a), "IDA's Performance-Based Allocation System: IDA Rating Disclosure and Fine-tuning the Governance Factor" http://siteresources.worldbank.org/IDA/Resources/PBAFINAL.pdf Accessed 23 October 2004.

World Bank (2004b), *Economic Growth in the 1990s: Learning from a Decade of Policy Reform,* Washington, D.C.

Comment

by
Serge Michaïlof
Director of Operations
Agence française de développement

First, I should like to thank Ravi Kanbur for the clarity of his analysis and the lucidity of his critiques—which leave but little space for the commentator—as well as the modesty of his proposals, which testify to his experience with large bureaucracies and his familiarity with the difficulties that outside actors face in dealing them. I would also like to give the modest viewpoint of the junior officer that I once was in the great fight against poverty at the World Bank: a "view from the trenches" of the CPIA-based aid allocation formulas. In the second part of my commentary, I will take this opportunity to describe my approach as director of operations at AFD, a small aid institution that is directly confronted with the issue of optimal resource allocation in a context where our subsidised resources are extremely scarce.

So thanks again to Ravi for his crystal-clear analysis. Thanks to his teaching talents and the illuminating simplicity of his expression, I think I have finally understood the logic behind the famous allocation of IDA resources. I must confess that this logic had been a mystery to me during all my years as a World Bank country director. Despite my intellectual laziness, I had, of course, understood that the allocation procedure obeyed set rules. Yet, I found some results incomprehensible, often counter-intuitive, and I was convinced that the process relied more on magic than on rational, logical analysis. First of all, the countries I dealt with never received the level of allocation that I felt was desirable or logical. Some countries—as it seemed with my reduced visibility from the bottom of my trench—received too much and were wasteful. Others did not receive enough, or the aid was not sufficiently timely for the corresponding resources to "make

a difference". I concluded, therefore, that malicious spirits were deforming some of "my" countries' PR indexes overnight so as to give the advantage to countries dealt with by "my colleagues", as they were obviously better at communicating with these malicious beings, I often had the impression, to use a farming metaphor, that my IDA allocations let me water dying or even dead crops abundantly, but they did not allow me to adequately water the most promising crops at the right moment in their growth cycle.

For a while, I suspected that a vast conspiracy favoured certain countries at the expense of others, for reasons that escaped me. I even imagined, dare I say it, that a bureaucratic plot was organising and rationalising, in pseudo-scientific guise, a simple system of "reputation-based grading" with broad categories, such as economic management, structural policies, etc. I wondered to what extent these "reputation-based gradings" were not based on old clichés or out-of-touch images, rather than on rigorous, up-to-date factual analysis; on the quality of the red carpet rolled out to welcome our mission, or even on the virtuosity of the accompanying violins. Ah, the long drawn-out sobbing of violins provided by certain heads of state who were particularly skilful with our officials! Was this not instrumental in upping performance by two or three points for certain criteria…? What other explanation for some of the more surprising rankings? Happily, as in all magic, is there not an invisible hand there to repair injustices? What else could explain why some French-speaking African countries are positioned in the top quintile, despite their official language and—especially when you examine the list carefully—their well-known performance regarding corruption and governance?

Yet, all in all, this bureaucratic magic fortunately makes it possible to rank countries and distribute aid. To be frank, this approach is unavoidable for any large aid institution like the World Bank, since an attempt at rational analysis, even with its series of subjective "reputation-based gradings", is better than nothing at all. And, fortunately, this magic works: Zimbabwe does not receive the same treatment as Cape Verde. Does this mean, however, that we should believe this approach is strictly rigorous?

In Section 3 of his paper, Ravi Kanbur shows, more aptly than I could ever do, its main conceptual weaknesses. My main focus is on the point he underlines about applying the same CPIA formula to all countries. This practice is underpinned by the belief that the much sought-after objective of growth is to be attained though a single

model of "best practice". This is still the case, despite the lessons to be learned from the now reputed work, *The Elusive Quest for Growth,* by our colleague, Bill Easterly.

I will give some nuance to Ravi's remark on how aid flows fail to influence policies. Coming back for a moment to his noteworthy, yet slightly over-hasty phrase: "*because if the experience of two decades has taught us anything, it is that the development assistance tail cannot wag the domestic political economy dog. Rather we should take the policies as emerging out of the domestic political economy*". This magnificent sentence rightly breaks up a certain idea of the effectiveness of "policy lending" conditionalities aimed at "buying" reforms. Ravi is perfectly right: you cannot bludgeon a donkey into drinking if it is not thirsty, any more than you can use conditionalities to bludgeon the way to structural reform. On the other hand, policy lending and conditionalities are powerful tools for change when conditions are favourable to reform in terms of political economy; i.e. when there is a real local will for reform (this happens, even in the most problematic, lowest-quintile countries). Not to forget that these reforming wills most often come up against political, institutional and other obstacles. Identifying such situations implies a political analysis that cannot be summed up in a series of CPIA-type criteria; it requires judgement that cannot be summarised by simple grading. Post-conflict situations are an extreme case in point, while LICUS (low-income countries under stress) countries are more generalised cases. In these particular circumstances, supporting the country's efforts demands resource levels at very short notice, levels which are far removed from the "normal" equation-based aid allocation, no matter how sophisticated the calculations.

This problem did not escape Ravi and he is rightly sceptical of the over-used "new dawn" concept. I can only approve of his mistrust, given the great many freshly appointed country directors I saw at the World Bank using the new dawn argument to negotiate exceptional IDA allocations for their countries. Scepticism, however, does not mean exclusion. Political opportunities do indeed exist and, in certain circumstances, there are indeed officials intent on carrying through courageous reforms. In fact, what needs to be criticised here is the viscosity of an approach that allocates resources primarily according to a collective perception of past performance, rather than assessing opportunities for change; the inertia of an approach that is out of step with the donors' *occasional* need for rapid shifts in stance. Certainly, Ravi's proposal of adding outcome criteria would attenuate the notion of "one size fits all", since the monitored outcomes

could depend on each country's development priorities, thus attenuating the conditionalities tied to a prescribed model.

The CPIA and the logical aid allocation formulas remind me of the optimised savings formulas that private investment funds propose you. These typically seek to optimise your income, conditional on risk-sharing, in a context of prudence and scepticism. The result is a "prudential"-style management of aid resources, and frequent irrigation of crops that are either dying or already dead. Development, however, happens in bursts. Its working principle is more like venture capital, than prudential funds. For the sake of efficiency, particularly in less well-performing countries, decisions on aid allocation must have the capacity (capacity is the word, since nothing should be systematic) to react to opportunities based on political judgements about the capacity of a government team, its seriousness, its desire to change and its courage.

I also sadly recall one Prime Minister with clear ideas and a firm will being appointed in a country that had abysmal governance and huge financial needs. I remember his disappointment at the heaviness of the Bank's procedures and the meagre IDA envelope (yet scientifically determined), when I told him the figure. I have kept the thank-you letter he sent me a year later after being expelled from office; in it he thanks me for my efforts to try to help him with the task. I deeply regret the fact that I was not able to take a *risk* on him based on the favourable judgement I had made… and the idea that a certain sum could have *perhaps* made a difference on the effectiveness of his action and for his country's future. As a good bureaucrat, I remained trapped by the magic formula.

In this light, I agree with Ravi's proposal to earmark funds (although perhaps not as little as he proposes) which would allow us to respond to such situations and to take "risks" on a team, backed by our own judgements of them, whilst awaiting performance-based results to take up the torch in a traditional process. In this context, his proposal to introduce a new criterion into the CPIA, based on effective results in terms of the MDGs, appears to be very sensible. It could help to improve the credibility of a process that is rather complex and logical, but not lacking its own weaknesses and subjectivity. The introduction of outcome criteria improves the relevance of what we are judging, i.e. the results rather than intermediary indicators (founded on the assumption that "one size fits all") which assess development processes and the probabilities of reaching these

results. However, this introduction does not alter the difficulties linked to a focus on the past (see supra).

Given my slight scepticism—which you have surely noticed by now—towards these mechanistic aid allocation processes, which are nonetheless indispensable to large donor, you are probably waiting for me to answer the question: how is it done at AFD?

I admit frankly that my approach in this domain is more akin to cooking than science, although there is nothing derogatory in the term "cooking", which is an art itself and an indispensable one to boot. As director of operations and thus "head cook", I am of course ready to reveal *my* secret recipes.

To begin with, there are two elements that differentiate our aid allocation issue from that of a large, multilateral allocation organisation.

First, there are historical and diplomatic factors which define countries eligible for AFD aid. Until recently, few countries were eligible for subsidies (in the ZSP, Solidarity Priority Zone) and, among them, priority is given to French-speaking countries. These historical and diplomatic factors—mentioned by Ravi—are not part of his study because his focus is on IDA. They are, however, determining parameters for all the important bilateral donors (the Netherlands, the DFID, etc.). The question of optimal allocation for a bilateral agency such as AFD is thus posed necessarily within a pre-defined area. Whether or not to take them into account in allocation formulas is, by the way, a worthwhile question (for example: introducing a specific variable that adds "value" to the French-speaking criterion in the equation). This first element is of a high significance, and can be summarised by the idea of aid allocation that occurs in a two-step process. This two-step evaluation completely changes the evaluation of resource allocation, in terms of aid agency performance. It also explains some results from comparative studies that judge the quality of this allocation based on a geographic perimeter, which is not relevant for a bilateral agency.

The second element to be taken into account from the outset is AFD's small size in foreign states. Here, the aid involved is less than a billion euros in annual commitments, excluding structural adjustment operations, which depend on political decisions and C2D operations which represent the French participation in the HIPC initiative. This small

sum is mainly split between aid in the form of loans with variable concessionality at more than 80%, and in the form of subsidy for the rest. However, the allocation rules for these resources vary according to the type of instrument and the geography concerned. For aid in the form of non-sovereign loans granted to sufficiently creditworthy public entities, we favour Sub-Saharan Africa. The recipe that I followed here implies, as is often the case with the culinary arts, both great rigour and flexibility.

Great rigour is necessary in three areas:

1) The choice of sectors that are eligible for these loans (infrastructure, local governments, social and environmental sectors);
2) Draconian rules for using concessionality (no crowding-out or "stroke of luck" effect; measurement and "negotiation" of this concessionality, which implies precise measures and realisations);
3) Sustainable levels of risk.

Great flexibility, on the other hand, is required in geographic choice:

In this area, we use—and I am rather embarrassed to say it in public—the principles of a rabbit hunt. We shoot at anything that does not have mixomatosis. So few rabbits (i.e. projects corresponding to our areas of expertise) are unaffected by mixomatosis (i.e. sound and bankable), that we are happy to participate in the financing of all the viable non-sovereign projects that have a significant impact on growth or on the MDGs in Sub-Saharan Africa (SSA). I have been proceeding like this for just over three years, and I have unfortunately had many more problems identifying creditworthy counterparts and bankable projects in our areas of expertise (for non-sovereign projects in SSA) than problems of optimal allocation of resources implying the possibility of choice and arbitrage. In fact, distinguishing programmable aid and non-sovereign operations (which are, essentially, non-programmable) is indispensable in such cases.

As for aid in the form of sovereign loans, the successive debt cancellations have left us with an even more restricted list of potential clients. A few of them are in SSA, but they are increasingly often emerging countries in the Mediterranean Basin and Southeast Asia. The restricted number of client countries leads us to (i) define objectives in terms of medium-term maximum level of outstanding debt by country according to

our ratio of risk weighted with a detailed analysis of country risk. These objectives determine the maximum amounts of annual commitment by country, according to hypothesis on the pace of disbursements and reimbursements; and (ii) to define country strategies that are clear, well targeted, and comprehensible according to our comparative advantages (iii) using these strategies to identify operations with the most significant impacts.

The limited number of sovereign loans that we make each year (maximum 20) allows for pragmatic management, under the constraint of an annual per-country ceiling regarding commitments. This method is not fully satisfying intellectually and should be refined progressively as we develop our activity in new countries. We will have to make choices regarding the use of subsidised resources, which constitute the (scarce) element of concessionality for our aid.

The real challenge for us is the rational allocation of our weak subsidised resources (around 175 million euros), although these choices cannot be dissociated from political decisions that favour certain regions: French-speaking Sub-Saharan Africa and Madagascar, Laos and Cambodia, the Palestinian territories and recently Afghanistan.

We are aware that our modest resources do not allow our financing to have a major influence in terms of *direct* impact on development (i.e. in terms of kilometres of roads built, number of children sent to school, etc.). This financing, however, can have a major impact on the development *process*, on the absorption capacity of countries, by facilitating reforms of sectoral policy, reform of institutions, creation of institutions, the definition of methods and procedures, and by bringing them the corresponding expertise.

The projects and programmes that we finance focus, therefore, on the critical aspects of the development process. They bring expertise and methodological progress that clear the paths for other donors or provide better raising of local resources.

In the war on poverty, our approach is thus similar to that of commando teams that focus their strikes on weak sensitive areas. In this perspective, in terms of allocation, volume distribution is, paradoxically, relatively secondary. What matters most is identifying these weak points and sensitive areas, concentrating our means on them, and using

our effectiveness to "treat" them—i.e. proposing sustainable solutions. Our approach is thus hostile to spreading aid thinly. I am convinced that an allocation of our modest subsidised resources based on an IDA-type model would mean a thin spreading of AFD aid.

Let us note that our approach is also absolutely hostile to the lighthouse and beacon syndromes, which would determine our sectors of intervention only according to our expertise. Our approach is definitely based on a concentration of our weak means on critical issues of the development process.

In this context, the countries eligible for subsidies are divided into four categories: from countries of category 1, where we hope to be present on all major issues, to the countries of category 4, where we do not wish to operate. This division into four categories is undertaken by crossing two series of criteria:

(i) the historical and diplomatic criteria mentioned earlier, which lead to a favouring of the French-speaking SSA, the Indochina Peninsula and some countries in crisis;

(ii) within this determined space, effectiveness criteria relating to a country's economic management and governance. This judgement is based on knowledge of the country, our own analysis (macroeconomic, sectoral, etc.) and/or analysis carried out by other donors, including the World Bank… It is not "scientifically" reduced to an equation.

An allocation is then made by large geographic region (western, central or southern Africa, Asia…) according to demographic considerations and GDP, and the number of countries of category 1, 2, 3, and 4 in each regional block. These approaches imply value judgements, and they are later validated or amended by the supervisory ministers.

Our regional directors then have a great liberty to allocate these subsidised resources according to clear and targeted country strategies, which are discussed at the level of our executive management. In a near future, these strategies will be part of the DCP (policy framework paper) prepared under the responsibility of the French ambassadors. Our concern is still to avoid spreading, to maximise our impact in terms of local absorption capacity building and to treat the critical points. According to Jean-Michel

Severino's favourite expression, we are thus attempting to "*weigh more than our financing*", even if the weight of this financing is far from negligible. To cite a concrete example: through our micro-financing operations, we reach a population of more than two million. Our real satisfaction, however, comes from setting up *X* number of successful institutions, which then constitute just as many "best practice" models… from implementing regulation frameworks adapted in *Y* number of countries… from having helped instruct *Z* number of specialists in this domain.

This approach requires reactivity, knowledge of the field and adaptability. Given the very low level of our current subsidies, this approach is not, in my opinion, compatible with a mechanical method of resource allocation, such as one determined by a mathematical formula. If our subsidised resources were multiplied by three— something I really hope for—, I would be ready to eat my hat and put priority on setting up a non-rigid allocation model. One which does not lead to spreading thin, but leaves room for political decision to refine the countries' envelopes.

In order to be effective, this approach demands integrating institutional and sector dimensions into our operational approach. Recently, in France, certain proposals have been put forward that aim at dissociating the operational activity from reflection, dialogue and support of sector policy. Such an approach badly misjudges the basis of aid effectiveness in a bilateral institution such as AFD. It would lead to splitting up light battalions into two; it would provoke a division of forces and a sector duplication. It would run contrary to all good aid practices. Development is fragile and complex enough without having donors uselessly complicate their procedures.

In all this, I have kept in mind the vivid memory of a remark Ravi made to me almost ten years ago, when we were both drafting a memorandum for a newly-appointed Vice-President of the World Bank. Ravi told me that in development, "*ideas are more important than money*". I have attempted to apply this principle in my work at AFD. And I acknowledge that I have not succeeded in finding a formula for the optimal allocation of good ideas between countries, although this is an exceptionally rare and precious factor.

Comment

by
Mark McGillivray*
WIDER, Helsinki, Finland

1. Introduction

Ravi Kanbur's paper is a very useful contribution to the literature on aid. It provides an incisive description and analysis of the IDA allocation procedure and its links to the World Bank's Country Policy and Institutional Assessment (CPIA). The paper (hereinafter referred to as Kanbur, 2004) provides a compelling case for the moving towards a development outcomes-based approach to aid allocation.

This comment looks at the fundamental conclusion of Kanbur (2004), that the IDA allocation formula ought to be augmented with additional, development outcome variables. The paper argues that, while the empirical literature is almost unanimous in the finding that growth would be lower (and as a consequence poverty levels higher) in the absence of aid, there remains some debate over the relevance of policy for these outcomes. Indeed, the clear majority of studies are consistent with the view that recipient country policy regimes do not systematically impact on aid effectiveness. Taken at face value, this would do more than justify a modest revision of the IDA allocation formula. Instead, it would appear to support a re-examination of the role of policy for aid effectiveness and possibly, on the basis of this re-examination, more substantive modifications than proposed by Kanbur (2004).

*. The author is a Senior Research Fellow and Project Director at the World Institute for Development Economics Research (WIDER) of the United Nations University in Helsinki, Finland. He is also an HDCA Fellow of Global Equity Initiative at Harvard University and an External Fellow of the Centre for Research in Economic Development and International Trade at the University of Nottingham.

2. Aid and policy

As pointed out in Kanbur (2004), the current IDA allocation formula, outlined in World Bank (2003), is as follows:

$$\frac{A_i}{P_i} = f\left[PR_i^{2.0}, GNIPC_i^{-0.125}\right] \qquad (1)$$

where A_i is aid to country i, P_i is i's population, PR_i is that country's performance rating and $GNIPC_i$ is its income per capita. Aid to any country i is an increasing function of the performance rating and a decreasing function of income per capita, according to this formula. The performance index rating is obtained from the World Bank's Country Policy and Institutional Assessment (CPIA) and Annual Review of Portfolio Performance (ARPP) and a governance factor, obtained largely from components of the CPIA, as described in Kanbur (2004). Kanbur correctly observes that the CPIA components are "central building blocks" in the construction of the performance rating (Kanbur, 2004, p. 8).

One way of assessing the appropriateness of the IDA formula is to look at research on aid effectiveness. There is now a large and growing literature on this subject, as surveyed by a number of recent studies (see Beynon, 2001, 2002, Hansen and Tarp, 2001, Kanbur, 2003, McGillivray, 2003a, 2003b, 2004 and Morrissey, 2001). Much of the recent (post-1997) literature has its origins in the seminal, yet still rather contentious, research of Burnside and Dollar (1997a, 2000). The econometric model estimated by Burnside and Dollar may be described by the growth equation outlined in Kanbur (2004):

$$G_i = \alpha + \beta Y_i + \theta X_i + \gamma A_i + \eta X_i A_i + \varepsilon_i \qquad (2)$$

where G_i is a measure of economic growth in country i, Y_i is a vector of country i structural variables, X_i is vector of variables describing policies in i and ϵ_i is a classical stochastic error term capturing variation in G_i not captured by the preceding variables.

A key and contentious finding of Burnside and Dollar (1997a, 2000) was that the impact of aid on growth could only be observed through its interaction with recipient

policies. Put differently, it was found that aid alone was ineffective and that its effectiveness was contingent on the policy regimes of recipients.[1] This was based on the finding that λ was insignificantly different from zero and that η was positive.

A number of subsequent studies have drawn the same conclusions for aid as Burnside and Dollar (1997a, 2000). All emanating from the World Bank, this group of studies consists of Burnside and Dollar (2004), Collier and Dollar (2001, 2002), Collier and Dehn (2001) and Collier and Hoeffler (2002). But there is far more evidence to suggest that aid works in countries irrespective of the quality of policy regime (Amavilah, 1998, Durbarry *et al.*, 1998, Svensson (1999), Hansen and Tarp (2000, 2001), Lensink and Morrissey (2000), Lensink and White (2001), Dalgaard and Hansen (2001), Guillamont and Chauvet (2001), Hudson and Mosley (2001), Lloyd *et al.* (2001), Lu and Ram (2001), Chauvet and Guillamont (2002), Dalgaard *et al.* (2004), Gounder (2001, 2002), Gomanee *et al. (*2002a, 2003), Ram (2003, 2004), Economides *et al.* (2004), Feeny (2004), Clemens *at al.* (2004), Heady *et al*. (2004), Outtara and Strobl (2004), and Roodman (2004). Some of these studies explicitly test for the relevance of policy, finding that aid's impact on growth is not contingent on the recipient policy regime. Others do not take this issue into account, but still find that aid and growth are positively associated.[2]

1. A parallel claim based on these results was that aid only works in countries with good policies (World Bank, 1998, Burnside and Dollar, 1997b). It is not clear that the Burnside and Dollar (1997a, 2000) results support such a conclusion. The consistent conclusion, it seems, is that while aid works in all countries, it works better in countries with better policy regimes

2. Irrespective of whether policy is important for aid effectiveness, it must be emphasised that both groups of studies agree that aid works, in one way or another. The evidence of this is now overwhelming. To the author's knowledge at the time of writing this paper, 35 empirical aid-growth studies have been conducted since Burnside and Dollar (1997a). Thirty-four of these studies provide original empirical results, obtained from either new or updated data sets, similar data sets but employing different empirical methods or both. The only study that does not provide new results is Collier and Dollar (2002), as it re-reports results shown in Collier and Dollar (2001). Including Burnside and Dollar (1997a), 36 studies were conducted during 1997 to 2004, therefore. Thirty-four of these studies conclude that aid works. The two papers that do not draw this conclusion are Easterly *et al.* (2003) and Jensen and Paldham (2003). Note, though, that the former study was concerned with critiquing Burnside and Dollar's econometric analysis, and not with the effects of aid *per se*. In a similar vein, the Roodman (2004) study shows that the results obtained by some studies can be fragile, and for this reason qualifies the results it reports. Further note that Ouattara and Strobl (2004) conclude that project aid worked but programme aid did not and Ram (2004) concludes while the overall impact of aid is positive, that multilateral unlike bilateral aid does not work. All the studies cited above were concerned specifically with the impact of aid on per capita national income growth, the exception being Lloyd *et al.*, which looked at private consumption growth. For a discussion of a range of related issues, see McGillivray (2003a, 2003b,) Collier and Dollar (2004) and Lensink and White (2000). Clemens *et al* (2004) cite number of additional studies, the overwhelming majority of which conclude that growth would be lower in the absence of aid.

This has profound implications for the IDA allocation formula. As Kanbur (2004) points out, from (2) it follows that the mathematical relationship between policies and aid effectiveness is:

$$\frac{\partial G_i}{\partial A_i} = \gamma + \eta X_i \qquad (3)$$

The clear majority finding is that the parameter _ takes a value of zero, or one that is not significantly different from zero statistically. It could be the case that the first group of studies mentioned above are methodologically superior to the second or, more generally, that each of the studies in the second group have got it wrong. The latter is unlikely, based on the sheer weight of numbers. The former is unlikely as well, as Easterly *et al.* (2004), Jensen and Paldham (2003) and Roodman (2004) show evidence that casts much doubt on the robustness of the Burnside and Dollar (1997a, 2000). Roodman also suggests that the Collier and Dehn (2001) and Collier and Hoeffler (2002) results are not so robust, and that Hansen and Tarp (2001) results are the most robust of a range of aid-growth studies.

3. To Augment or... ?

What are the implications of these findings for aid allocation? Should one simply ignore the role of policies, allocating aid to countries irrespective of their policy regimes? Is aid equally effective in all countries, no matter how good or bad their policies might be? One would, in principle, expect that policies will matter, in one way or another, potentially or otherwise. So a reasonable answer to these questions is 'no'. Possibly reflecting this, there is a general acceptance among researchers that better policies, however defined, should in all probability result in more effective aid.[3] This is based on theoretical reasoning, evidence from micro studies and field experience. What is clear, though, is that the clear majority of studies reject the specific relationship depicted by equation (3) and, given this, do not support the *mathematical* basis of the IDA formula.

This, in turn, has implications for the main conclusion of Kanbur (2004), that one should augment the IDA formula by adding outcome variables to vector of policy

3. See Tarp (2000), Benyon (2001, 2002), Gunning (2001), Morrissey (2001), McGillivray (2003a, b) and Collier and Dollar (2004).

variables, X_j. It might be the case that the absence of such variables is why a clearer link between policies and aid effectiveness has not been established. However, one would reasonably expect that there is more to the story than the absence of some variables. The link between policies and aid effectiveness is, in all probability, far more complex than depicted by equation (2). Threshold effects, non-linearity, endogeneity and a range of other complex factors are likely to play some role.[4] What would appear essential is greater knowledge of these factors. Rather than augmenting the existing formulae, this might justify more profound changes to the processes through which aid is allocated among recipient countries.

4. Conclusion

Kanbur (2004) provides a clear and compelling case for revising the IDA allocation formula. The proposed revisions are modest. Such revisions are better than none at all and are also more likely to be implemented than larger ones. This comment argued, however, that what is required more than anything else is a better knowledge of what makes aid work, and that revisions to aid allocation formulae should be considered in this light. This may justify more radical changes to the IDA formula than those outlined in Kanbur (2004).

REFERENCES

Amavilah, V.H. (1998), "German Aid and Trade versus Namibian GDP and Labour Productivity", *Applied Economics*, Vol. 30, pp. 689-695.

Burnside, C. and D. Dollar (1997a), *Aid, Policies, and Growth*, Policy Research Working Paper No. 1777, World Bank, Washington DC.

Burnside, C. and D. Dollar (1997b), "Aid Spurs Growth – in a Sound Policy Environment", *Finance and Development*, December, pp. 4-7.

4. Equally, there are likely to be a range of other factors, quite independent of policies, that will impact on aid effectiveness.

Burnside, C. and D. Dollar (2000), "Aid, Policies, and Growth", *American Economic Review*, Vol. 90, No.4, pp. 847-868.

Burnside, C. and D. Dollar (2004), *Aid, Policies and Growth: Revisiting the Evidence*, World Bank Policy Research Paper No. O-2834, March, World Bank, Washington D.C.

Chauvet, L. and P. Guillaumont (2002), "Aid and Growth Revisited: Policy, Economic Vulnerability and Political Instability", Paper presented at the Annual Bank Conference or Development Economics on Towards Pro-poor Policies, June, Oslo.

Clemens, M. and S. Radelet (2003), *The Millennium Challenge Account: How Much is Too Much, How Long is Long Enough?*, Centre for Global Development Working Paper No. 23, Centre for Global Development, Washington.

Clemens, M., S. Radelet and R. Bhavnani (2004), *Counting Chickens when They Hatch: The Short-term Effect of Aid on Growth*, Centre for Global Development Working Paper No. 44, Centre for Global Development, Washington.

Collier, P. and J. Dehn (2001), *Aid, Shocks, and Growth*, World Bank Policy Research Working Paper No. 2688, World Bank, Washington.

Collier, P. and D. Dollar (2001), "Can the World Cut Poverty in Half? How Policy Reform and Effective Aid Can Meet the International Development Goals?", *World Development*, Vol. 29, No. 11, pp. 1787-1802.

Collier, P. and D. Dollar (2002), "Aid Allocation and Poverty Reduction", *European Economic Review*, Vol. 26, pp. 1475-1500.

Collier, P. and D. Dollar (2004), "Development Effectiveness: What Have We Learnt?", *Economic Journal*, Vol. 114, No. 496, pp. F244-F271.

Collier, P. and A. Hoeffler (2002), *Aid, Policy, and Growth in Post-Conflict Societies*, Policy Research Working Paper No. 2902, World Bank, Washington.

Dalgaard, C. and H. Hansen (2001), "On Aid, Growth and Good Policies", *Journal of Development Studies*, Vol. 37, No.6, pp. 17-35.

Dalgaard, C., H. Hansen and F. Tarp (2004), "On the Empirics of Foreign Aid and Growth", *Economic Journal*, Vol. 114, No. 496, pp. F191-F216.

Durbarry, R., N. Gemmell and D. Greenaway (1998), *New Evidence on the Impact of Foreign Aid on Economic Growth*. CREDIT Research Paper No. 98/9, Centre for Research in Economic Development and International Trade, University of Nottingham, Nottingham.

Easterly, W., R. Levine and D. Roodman (2004), "New Data, New Doubts: A Comment on Burnside and Dollar's 'Aid, Policies and Growth' (2000)", *American Economic Review*, Vol. 94, No. 3.

Economides, G., S. Kalyvitis and A. Philippopoulos (2004), "Do Foreign Aid Transfers Distort Incentives and Hurt Growth? Theory and Evidence from 75 Aid-recipient Countries", Athens University of Economics and Business, mimeo.

Feeny, S (2004), "The Impact of Aid on Economic Growth in Papua New Guinea", *Journal of Development Studies*, forthcoming.

Gomanee, K., S. Girma and O. Morrissey (2002), *Aid and Growth in Sub-Saharan Africa: Accounting for Transmission Mechanisms*, CREDIT Research Paper 02/05, Centre for Research in Economic Development and International Trade, University of Nottingham, Nottingham.

Gomanee, K, S. Girma and O. Morrissey (2003), *Searching for Aid Threshold Effects: Aid, Growth and the Welfare of the Poor*, CREDIT Research Paper 02/05, Centre for Research in Economic Development and International Trade, University of Nottingham, Nottingham.

Gounder, R. (2001), "Aid-Growth Nexus: Empirical Evidence from Fiji", *Applied Economics*, Vol. 33, pp.1009-1019.

Gounder, R. (2002), "Empirical Evidence of the Relationship between Foreign Aid and Economic Growth: The Case of the Solomon Islands", in B. Mak Arvin (ed.), *New Perspectives on Foreign Aid and Economic Development*, Praeger, Westport.

Guillaumont, P. and L. Chauvet (2001), "Aid and Performance: A Reassessment", *Journal of Development Studies*, Vol. 37, No. 6, pp. 66-87.

Gunning, J.W. (2001), "Rethinking Aid" in B. Pleskovic and N. Stern (eds.), *Annual World Bank Conference on Development Economics 2000*, World Bank, Washington.

Hansen, H. and F. Tarp (2000), "Aid Effectiveness Disputed", *Journal of International Development*, Vol. 12, pp. 375-398.

Hansen, H. and F. Tarp (2001), "Aid and Growth Regressions", *Journal of Development Economics*, Vol. 64, No. 2, pp. 547-570.

Heady, D, D. Prasad Rao and A. Duhs (2004), "All the Conditions for Effective Foreign Aid", University of Queensland, mimeo.

Hudson, J. and P. Mosley (2001), "Aid, Policies and Growth: In Search of the Holy Grail", *Journal of International Development*, Vol. 13, pp. 1023-1038.

Islam, N. (2002), "Regime Changes, Economic Policies and the Effects of Aid on Growth", Paper presented at the Conference on Exchange Rates, Economic Integration and the International Economy, May, Ryerson University, Canada.

Jensen, P.S. and M.Paldam (2003), *Can the New Aid-Growth Models be Replicated?*, Working Paper No. 2003-17, Institute for Economics, Aarhus.

Kanbur, R. (2003), "The Economics of International Aid", http://www.arts.cornell.edu/poverty/kanbur/HandbookAid.pdf, Accessed October 23, 2004.

Kanbur, R. (2004), "Reforming the Formula: A Modest Proposal for Introducing Developmental Outcomes in IDA Allocation Procedures", paper presented at the 2nd AFD-EUDN Conference, Paris, November, 2004.

Lensink, R. and O. Morrissey (2000), "Aid Instability as a Measure of Uncertainty and the Positive Impact of Aid on Growth", *Journal of Development Studies*, Vol. 36, No. 3, pp. 31-49.

Lensink, R. and H. White (2000), "*Assessing Aid*: A Manifesto for the 21st Century?", *Oxford Development Studies*, Vol. 28, No. 1, pp. 5-18.

Lensink, R. and H. White (2001), "Are There Negative Returns to Aid?", *Journal of Development Studies*, Vol. 37, No. 6, pp. 42-64.

Lloyd, T., O. Morrissey and R. Osei (2001), *Aid, Exports and Growth in Ghana*, CREDIT Research Paper No. 01/01, Centre for Research in Economic Development and International Trade, University of Nottingham, Nottingham.

Lu, S. and R. Ram (2001), "Foreign Aid, Government Policies and Economic Growth: Further Evidence from Cross-country Panel Data for 1970 to 1993", *International Economics*, Vol. 54, pp. 15-29.

McGillivray, M (2003a), "Aid Effectiveness and Selectivity: Integrating Multiple Objectives in Aid Allocations", *DAC Journal*, Vol. 4, No. 3, pp. 23-36.

McGillivray, M. (2003b), "Efficacité de l'aide et sélectivité : vers un concept élargi" ("Aid Selectivity and Effectiveness: Towards a Broader Concept"), *Revue d'Économie du Développement*, No. 4, pp. 43-62.

Morrissey, O. (2001), "Does Aid Increase Growth?", *Progress in Development Studies*, Vol. 1, No. 1, pp. 37-50.

Ouattara, B. and E. Strobl (2004), *Disaggregating the Aid and Growth Relationship*, School of Economic Studies Discussion Paper No. 0414, University of Manchester, Manchester.

Ram, R. (2003), "Roles of Bilateral and Multilateral Aid in Economic Growth of Developing Countries", *Kyklos*, Vol. 56, pp. 95-110.

Ram, R. (2004), "Recipient Country's 'Policies' and the Effect of Foreign Aid on Economic Growth in Developing Countries: Additional Evidence", *Journal of International Development*, Vol. 16, pp. 201-211.

Roodman, D. (2004), *The Anarchy of Numbers: Aid, Development and Cross-country Empirics*, CGD Working Paper No. 32, July 2004, Centre for Global Development, Washington D.C.

Tarp, F. (ed.) (2000), *Foreign Aid and Development*, Routledge, London and New York

World Bank (1998), *Assessing Aid: What Works, What Doesn't and Why?*, Oxford University Press, New York.

World Bank (2003), "Allocating IDA Funds Based on Performance: Fourth Annual Report on IDA's Country Assessment and Allocation Process", http://siteresources.worldbank.org/IDA/Resources/PBAAR4.pdf, Accessed October 23, 2004.

Comment

by
Nicolas van de Walle
Jack Knight Professor of International Studies,
Director of the Mario Einaudi Center for International Studies
and Professor of Government, Cornell University

In his interesting contribution to this conference, Ravi Kanbur proposes that the World Bank change its aid allocation mechanism to take into account improvements in economic outcomes and more specifically the progress recipient countries make in reaching the Millennium Development Goals. Currently, the Bank's approach does not take into account such improvements and instead uses a set of policy and governance variables that the Bank believes to be related to economic outcomes. Kanbur suggests, and I agree, that it makes more sense to link allocation to indicators of actual performance, and thus not make the assumption that all low-income countries benefit in the same manner from the same policies and governance patterns.

In these brief comments, I will first develop several points relating directly to Prof. Kanbur's proposal. I will then broaden my perspective somewhat to examine several assumptions that are implicit in his argument.

First, Kanbur's essay is most useful as a primer on the principles with which the world's biggest development agency seeks to allocate its aid. I suspect many people outside of the Bank will be as surprised as I was to discover that the Bank has such a precisely specified system, that—as Kanbur makes clear—it is based on rather arbitrary weights, and that it weighs governance criteria as much as it appears to. It would indeed be interesting to find out more about the evolution of the Bank's allocation

process over the course of the last several decades, and when and how governance criteria emerged. As interesting, would be an empirical discussion of the impact of this system on actual Bank lending. Kanbur makes clear the Bank considers the allocation norm indicative rather than an entitlement, and that a number of other considerations help determine the level of aid provided to a country by the Bank. The Bank is not the only public organisation where principle and practice diverge, so the extent to which it conforms to its own norms would provide valuable information.

Kanbur's essay might also have provided some analysis of the extent to which the changes he proposes in the Bank's formula would alter its actual allocations. Even a small number of empirical examples, showing that adding an outcome variable would have a significant impact on the allocation to some current Bank clients, would have strengthened his line of argument considerably. While I agree with his presumption that there are multiple paths to development, it is hard to believe that the paths diverge so much that the type of outcome variables he proposes would not co-vary relatively closely with many of the existing CPIA criteria. My own nagging suspicion is that implementing Kanbur's "modest proposal" would in fact have little effect on current allocation patterns.

Second, Kanbur criticises the current allocation mechanism as based on the logic that all low-income countries will follow the same path to development. He points out that, even if this is true for the "average country", there is important variation across all Bank clients that reflect "specific local factors". I agree, but it seems to me there are political economy reasons for supporting a single formula for all countries: explicit rules eliminate discretion and prevent Bank staff from allocating aid resources for idiosyncratic and subjective reasons. Donors that do not follow a clear set of explicit common criteria for their aid are more likely to favour non-developmental objectives, and internal pressures to lend within aid agencies are more likely to lead to various rationalisations to provide more aid than is optimal. There is much evidence to suggest that donors often enter into privileged relationships with recipient country governments and exaggerate progress on development objectives to justify continued assistance. The explicit guidelines for lending allocations are thus useful as "agencies of restraint", by helping to prevent such situations.

Third, the resort to performance criteria for aid allocation is not fully explained by Kanbur. Why would donors care about policy and governance? For much of aid's

history, donors did not much care about these things, focusing either on need or on nondevelopmental factors, such as commercial, foreign policy and humanitarian considerations (van de Walle, 2001). Much aid still follows these other forms of logic. Kanbur suggests that donors care about the effect of aid resources. If aid promotes economic development only in countries that follow certain policies than you get more bang for your buck if you direct aid resources to "virtuous" countries. This is clearly true, though I believe it has more of an effect on allocation at the Bank than for most other donors, who are less likely to base allocation decisions on their interpretation of the quality of policy.

There is another consideration, however: in recent years, donors have justifiably become concerned that aid resources can create a moral hazard effect, in relation to bad policy: high levels of aid allow governments to sustain bad policies, even when the aid was motivated specifically to help eliminate them. Much of the aid community has come to agree that the slow pace of economic reform in low-income countries in the 1980s and 1990s was due to this moral hazard effect, as donor conditionality proved ineffectual. As Paul Collier famously pointed out, Kenya received five successive loans from the IFIs to undertake the same set of agricultural reforms, and each time reneged on its promise (Collier, 1997). Indeed, given this reality, I find it odd that Kanbur argues that aid has little or no effect on the local political economy. In countries in which two thirds of the governmental budget is financed from aid, it seems clear that donors do have a powerful impact on policies and local politics. This is no less true whether or not donors are aware of it.

In sum, another justification for selectivity in lending has been the need to overcome these moral hazard effects. The current logic is that governments now need to prove their commitment to sound policies before they can receive Bank lending. Selective lending is also designed to signal to recipient countries that the aid regime has changed and that recipients will now be held accountable for their actions. This logic has led donors such as the Bank to publicly emphasise policy and governance performance criteria in their allocation.

But can a single donor like the Bank change recipient government behaviour by changing its allocation criteria? My own view is that even an important donor cannot have such influence unless it coordinates its allocation with other donors. In fact, donor

coordination has progressed little in the last decades. Even if some (but by no means all) donors have increased the selectivity of their aid, it can still be the case that overall aid flows remain largely random, if different donors do not demonstrate the same preferences in their actual allocation of aid. Thus, the Bank's decision to lend more money to a low-income country which has demonstrated an improvement in its infant mortality rates may not send a clear signal in a country in which the Bank only provides a quarter of total aid, and other donors do not share the Bank's priorities. Perhaps another donor prefers to reward countries that support "the war on terrorism", while another one only gives aid to its ex-colonies. In this situation, the aid community sends no clear signal to recipients.

Donor agencies have typically devoted too much time to elaborating their own allocation decision-rules and too little time worrying about the logic shaping overall aid flows. Yet, surely it is the latter that matters for development outcomes in a world in which the typically low-income country receives aid from two dozen donor agencies. Kanbur's proposals, while undeniably judicious, would have gained traction by addressing the broader issues of donor coordination. These strike me as the key to more effective aid.

REFERENCES

Collier, P. (1997), "The Failure of Conditionality", in C. Gwin and J. M. Nelson (eds.), *Perspectives on Aid and Development*, Overseas Development Council Policy Essay Number 22, Johns Hopkins University Press for the ODC, Baltimore, pp. 51-78.

van de Walle, N. (2001), *African Economies and the Politics of Permanent Crisis, 1979-1999,* Cambridge University Press, New York.

Aid and Health

by
Clive Bell
South-Asia Institute, University of Heidelberg, Germany;
and Human Development Network,
The World Bank and Carsten Fink International Trade,
World Bank Institute, World Bank Geneva Office

1. Introduction

We are charged with the task of answering the general question: "How should aid be allocated in the sphere of health?" The economist's pat answer is that aid, like all public funds, should be allocated to those programmes and projects that yield a positive net present value when their inputs and outputs are calculated at shadow prices, that is to say, at prices that properly reflect social scarcities. This is fine as a statement of the principle, but the policy maker can be forgiven for finding it unhelpful. It tells us very little about where to look with the aim of setting general priorities in practice.

With this aim in view, a natural starting point is to identify the diseases that afflict the populations of less developed countries most heavily. For if efficient means to combat them are available or can be developed, then there is a fair prospect of allocating aid to good purpose. Table 1 presents an overview of the global burden of disease, using disability-adjusted life years (DALYs)[1] as a yardstick. Each entry is the sum of the years of life lost due to disability and premature mortality caused by the disease in question. Three categories—communicable diseases, maternal and perinatal conditions, and nutritional deficiencies—accounted for 42% of the global

1. This measure is defined and discussed in Section 3.3.

Table 1.
Burden of disease in DALYs by cause in different regions, 2001

	World		Africa		The Americas	
	in thousands	%	in thousands	%	in thousands	%
TOTAL DALYs	**1,467,257**	**100.0**	**357,883**	**100.0**	**145,217**	**100.0**
A. Communicable diseases, maternal and perinatal conditions and nutritional deficiencies	**615,737**	**42.0**	**261,456**	**73.1**	**27,116**	**18.7**
Infectious and parasitic diseases	359,377	24.5	189,047	52.8	12,555	8.6
Tuberculosis	36,040	2.5	8,941	2.5	945	0.7
STDs excluding HIV	12,404	0.8	5,141	1.4	635	0.4
HIV/AIDS	88,429	6.0	67,460	18.8	2,767	1.9
Diarrhoeal diseases	62,451	4.3	21,524	6.0	2,794	1.9
Childhood diseases	48,268	3.3	24,997	7.0	490	0.3
Meningitis	6,420	0.4	964	0.3	695	0.5
Hepatitis Bc	1,684	0.1	274	0.1	125	0.1
Hepatitis Cc	844	0.1	141	0.0	99	0.1
Malaria	42,280	2.9	36,012	10.1	108	0.1
Tropical diseases	12,994	0.9	6,251	1.7	805	0.6
Leprosy	177	0.0	16	0.0	18	0.0
Dengue	653	0.0	6	0.0	90	0.1
Japanese encephalitis	767	0.1	0	0.0	0	0.0
Trachoma	3,997	0.3	1,526	0.4	0	0.0
Intestinal nematode infections	4,706	0.3	674	0.2	624	0.4
Respiratory infections	94,037	6.4	29,873	8.3	3,528	2.4
Maternal conditions	30,943	2.1	11,328	3.2	1,843	1.3
Perinatal conditions	98,422	6.7	21,920	6.1	7,096	4.9
Nutritional deficiencies	32,958	2.2	9,288	2.6	2,093	1.4
B. Non-communicable conditions	**672,865**	**45.9**	**66,105**	**18.5**	**97,402**	**67.1**
C. Injuries	**178,656**	**12.2**	**30,323**	**8.5**	**20,699**	**14.3**

Note: Definition of regions based on WHO classification. Europe includes Central Asia; Easter Mediterranean includes Pakistan; South-East Asia includes India; Western Pacific includes Australia, Cambodia, China, Japan, Malaysia, New Zealand, the Philippines, Korea, Singapore, and Vietnam.
Source: World Health Organisation, *World Health Report 2002*, statistical annex.

Eastern Mediterranean in thousands	%	Europe in thousands	%	South East Asia in thousands	%	Western Pacific in thousands	%
136,221	**100.0**	**151,224**	**100.0**	**418,844**	**100.0**	**257,868**	**100.0**
67,137	**49.3**	**14,607**	**9.7**	**188,152**	**44.9**	**57,269**	**22.2**
34,741	25.5	5,876	3.9	93,995	22.4	23,163	9.0
2,989	2.2	1,679	1.1	15,968	3.8	5,518	2.1
1,323	1.0	361	0.2	4,317	1.0	627	0.2
1,713	1.3	908	0.6	13,608	3.2	1,973	0.8
10,784	7.9	831	0.5	22,378	5.3	4,141	1.6
7,189	5.3	394	0.3	12,959	3.1	2,238	0.9
1,073	0.8	441	0.3	2,585	0.6	662	0.3
172	0.1	90	0.1	613	0.1	410	0.2
84	0.1	75	0.0	255	0.1	190	0.1
2,050	1.5	20	0.0	3,680	0.9	409	0.2
1,055	0.8	8	0.0	4,391	1.0	484	0.2
16	0.0	0	0.0	119	0.0	7	0.0
85	0.1	0	0.0	360	0.1	112	0.0
81	0.1	0	0.0	347	0.1	340	0.1
603	0.4	0	0.0	248	0.1	1,621	0.6
267	0.2	8	0.0	1,550	0.4	1,582	0.6
11,729	8.6	3,626	2.4	32,904	7.9	12,378	4.8
4,129	3.0	751	0.5	10,027	2.4	2,864	1.1
12,463	9.1	2,866	1.9	39,495	9.4	14,581	5.7
4,075	3.0	1,489	1.0	11,731	2.8	4,283	1.7
52,611	**38.6**	**115,902**	**76.6**	**176,569**	**42.2**	**164,276**	**63.7**
16,399	**12.0**	**20,714**	**13.7**	**54,123**	**12.9**	**36,323**	**14.1**

disease burden, non-communicable diseases for 46%, and injuries for the remaining 12%. A striking feature of the global picture is that the burden of communicable diseases is far higher in less developed countries than in rich ones. Indeed, communicable and related diseases account for almost three-quarters of the total burden of disease in Africa, and 45 to 50% in the Eastern Mediterranean and South and Southeastern Asia. More detailed examinations of the data (Gwatkin and Guillot, 2000; Bonilla-Chacin and Hammer, 2003) reveal that most of the difference in mortality between rich and poor countries and between rich and poor people generally arises from the incidence of the three named disease categories.

It is on these diseases that we shall concentrate in this paper. Table 1 demonstrates their importance in influencing the lives and welfare of the populations of developing countries. There is, however, another general, compelling reason to devote particular attention to communicable diseases, namely, their "public" nature, both within and across national boundaries. It is precisely the international character of these "flows" that strengthens the case for directing international flows of public resources to combat them. By restricting our treatment of the topic in this way, we shall also keep the scope of the paper within reasonable bounds.

Given the restriction to communicable diseases, we address the following questions:

- What principles and criteria should be employed in allocating aid flows for this purpose? In particular, how should the preservation of existing human capital and the formation of new human capital enter into the reckoning?
- What do these criteria imply for the ranking of diseases and the importance of prevention and treatment relative to R&D?
- What special considerations apply to the allocation of aid and the design of related incentive mechanisms aimed at promoting R&D on the main communicable diseases?

Even so, the paper cannot pretend to be comprehensive. In particular, we make no attempt to assess the performance of existing, and the likely pay-offs to prospective, aid programmes, including those aimed at realising the so-called 'Millennium Development Goals' (MDGs). We do, however, provide specific examples to illustrate and buttress our arguments.

The plan of the paper is as follows. Section 2 sets out a more detailed justification for directing aid especially towards activities that combat communicable diseases, as opposed to supporting other activities in the health sector or elsewhere. Section 3 lays out the general framework, the three components of which are: (i) the costs of disease; (ii) prevention, treatment and R&D; and, (iii) measuring the pay-off to spending. We then discuss how to allocate aid, first, in the realm of prevention and treatment (Section 4), and second, in that of R&D (Section 5). We summarise our findings and make certain proposals in Section 6.

2. The justification for aid

The general ethical and economic justifications for giving aid need no repetition here. They do not, however, exhaust the list of compelling reasons to allocate aid to combating communicable diseases. The following four, specific reasons are especially weighty.

1. Communicable diseases are, in effect, a "common property" resource, so that the allocation of resources in this sphere is prone to market failure. Within national borders, it is evident that the workings of the internal political economy often hinder an adequate public response. When the problems of fungibility are not too severe, aid can therefore have a beneficial influence on the welfare of the citizens of the recipient country.

2. Many important communicable diseases recognise no national boundaries, so that aid is an appropriate instrument to deal with a global common "ill". Allocating aid for this purpose can be justified on the grounds of the donor's pure self-interest.

3. R&D produces outputs with a strong public good component. Private R&D activities in health are concentrated in developed countries and are very heavily directed towards the health problems of their populations; for it is there that willingness-to-pay manifests itself as effective demand in the global marketplace. Aid aimed at furthering R&D on the main communicable diseases in developing countries can alter effective demand in their favour. As in reasons 1 and 2, allocating aid for this particular purpose can be justified on both ethical grounds

and the desirability of correcting global market failures that arise from externalities.

4. Diseases such as AIDS and tuberculosis, which afflict mainly young adults, not only destroy existing human capital, but can also weaken the inter-generational mechanisms by which new human capital is created through childrearing in the family and formal education. How well these mechanisms function depends on insurance arrangements and the quality of childrearing when one or both natural parents die prematurely. Since the "markets" for these goods are either imperfect or non-existent, reductions in premature adult mortality not only preserve existing human capital, but also promote the formation of future human capital. For this reason, aid can generate potentially important, long-term positive externalities, as well as combating the negative externalities stemming from the existence of pools of diseases.

3. The framework

Given these justifications, how should aid in this sphere be allocated? In order to answer this question, we first need a framework for identifying and evaluating alternative allocations, however they be financed. In doing so, we must address the question of whether one can draw up a good, concerted plan of action simply by choosing some set of initiatives, or "projects", from a collection of proposals that happen to be "on the desk". The theory and practice of planning tell us, on the contrary, that arriving at a good "plan" requires that projects be sought out, or even developed from scratch, in particular sectors. This conclusion implies the need for some scheme to organise and classify activities in such a way that promising projects, or prospects for projects, can be more easily identified. In effect, the scheme should enable us to organise our ideas about the broad priorities among groups of activities and even to draw some conclusions at the very outset.

The specific projects that emerge from this process must then be subjected to rigorous cost-benefit analysis; the proposed elements yielded by this *ex ante* scheme must pass muster when judged by a common standard. In order to implement such a procedure, we need to establish the effects of expenditures on various measures to

combat particular diseases, and to choose a way of transforming these effects into a measure of the resulting social pay-off. In essence, this is just a summary of the standard approach to all social cost-benefit analyses, but formulated for the case at hand. If we can agree on this much, then we will have established both the principles for allocating aid expenditures aimed at combating infectious diseases and the procedures for applying them in practice.

When drawing up a plan to combat infectious diseases, there are two natural ways of classifying activities: first, by disease; and second, by the measures to deal with them, namely, prevention, treatment, and research and development. These are perfectly compatible classifications, though it is natural and sensible to discuss prevention, treatment, and research and development in relation to each disease. We take them up in that order.

3.1. The costs of disease[2]

A listing of diseases by itself reveals virtually nothing. What we need is the effects of each disease on age-specific morbidity and mortality, with their ensuing consequences for economic welfare, broadly defined. In order to see what is involved, we begin with two illustrative, extreme examples of a single episode of illness or injury. We then present the elements more formally and extend the discussion to cover the case where the disease is infectious.

The first is drawn from van Dillen's (2004) study of vulnerability in a village in Tamil Nadu, south India. A young agricultural labourer was butted by a bullock while ploughing early in the irrigation season, suffering serious injuries to his left hand and arm, followed by a fever with the ensuing infection. Three weeks later, having spent Rs. 500 (about 10 days' wages) on visits to the doctor and medicines, he was still in no condition to work. His wife already had an infant to care for, and having to look after him too further reduced her earnings from wage labour and rolling *beedis* (cigarettes). Despite a little help from relatives, their short-term resources were then exhausted, and they took out a loan of Rs. 1000 at 10% a month from his employer. The young man was able to resume full-time work after another month, but the debt was still outstanding some months later.

2. This subsection draws heavily on Section 3 of Bell and Lewis (2004).

The resulting costs of his injury are broadly of three types. First, his inability to work and illness resulted in a loss of family earnings and a heavier workload for his wife. Secondly, treating his injury made a direct claim on resources, in the form of medical expenses. Thirdly, there were what can be thought of as the intangibles—in this case, the misery caused by the pain, the anxiety about finances, and the loss of leisure among family and friends. Who actually bears these costs in any particular case—it should be emphasised that they must be borne by some party or other—depends on insurance arrangements. In van Dillen's example, as in most poor economies, coverage will be provided largely by the family itself—though treatment may contain an element of public subsidy. The intangible costs of illness and disability, the young man and his wife had to bear themselves.

Now consider, in contrast, a minor civil servant in his mid-30s with two small children, who suffers a heart attack so massive that he is unable to resume working at all before he succumbs to a second and fatal one. All of his future output, dating from the time of the first attack to expected retirement, with a suitable adjustment for "normal" morbidity and mortality during that span of life, is thereby lost. The treatment after the first attack, the efforts at rehabilitation and the final, vain effort to resuscitate him after the second attack are relatively expensive, particularly if he was lucky enough to be admitted into an intensive care unit. The only "saving" under this heading is the expected costs that would arise from the treatment and care of his ailments in old age, which he fails to reach, and the disbursement of social security earned at retirement. The costs of the intangibles are no less sobering. The physical pain is accompanied by sadness at the prospect of an early death and what that will mean for his spouse and children. For them, there is not only the grief at his death, but also the enduring loss of his love, companionship and particular contribution to family life and childrearing. His insurance arrangements are unlikely to cover the first two types of costs, and cannot, in the nature of things, fully deal with the last.

These examples illustrate the importance of the family as an institution for, among other things, the pooling of resources. If the lost earnings and the cost of medical treatment are not fully covered by insurance arrangements with third parties, then one individual's illness will directly and adversely affect the well-being of other members, particularly the children, whose formal education may well suffer. Indeed, in the second example, it is perhaps the children who stand to lose most, a conclusion

that applies with undiminished force if their mother were to die early. These inter-generational effects imply that premature mortality among parents does potential long-term damage that goes well beyond the calculations involving their own lost earnings and the costs of treating their final illness, regardless of whether that illness is an infectious one.

It will be useful to formulate the elements of the above discussion in a somewhat more general way. As a reference case, one can think of an "ideal" trajectory, along which the individual enjoys robust good health throughout an untroubled childhood, a long and productive working life, and a full old age, and then dies suddenly and peacefully in bed. Each episode of illness disturbs this trajectory in some measure, perhaps catastrophically. The episode may have no lasting effects; it may so damage the individual's constitution as to increase the chances of various illnesses in the future, possibly with a fatal outcome; or it may itself be fatal. Depending on its nature, it brings with it one or more of the immediate effects set out above, to which must be added the *expected* value of the future costs that will arise if the present episode increases the chances of illness in the future, as happens, for example, when a childhood illness arrests the development of an organ.

What additional considerations come into play if the disease is infectious? The immediate consequence is the risk that the sick individual will infect others, who will go on to propagate it in their turn. There is therefore a further chain of *expected* costs, whose magnitude depends not only on the damage the disease does to each individual, but also on the probability that it will be transmitted. So long as that probability is at all positive, the illness is not a purely private matter for the infected individual and those of the immediate family who share in a common economic fortune, as set out in the above examples. Rather, an infection has a certain *public* character in that the infected person can, wittingly or otherwise, transmit the disease to others without their consent or even without their knowledge. Any case of a communicable disease therefore generates, at least potentially, negative externalities. If transmission does occur, the parties contracting the infection will often become aware of the fact only after the event, and no doubt would have rejected the transaction had they been aware that it was "on offer" at the time. The very fact that the threat of transmission exists will itself cause economic costs to arise, depending on the extent to which measures are taken to ward it off.

Thus far, we have implicitly treated the aggregate costs of morbidity and mortality as the sum of the costs arising from individual cases, the timing and number of which depend on the prevalence of (infectious) diseases and the age structure of the population. This procedure is not fundamentally invalid if everything is correctly measured, but it does raise the question of how prices—and especially factor prices—depend on the levels of morbidity and mortality. In order to address this question, we suppose that there is some prevailing, "normal" pattern of morbidity and mortality in the population, which provides the necessary benchmark. A single, additional case of a disease causes a very small disturbance to this pattern, and hence will have a negligible effect upon current and future prices—unless it arouses the suspicion that many more cases will follow, a possibility we shall take up in a moment. An actual epidemic, in contrast, produces a measurable change in overall morbidity and mortality, however temporary. The associated effects on the structure of relative prices depend not only on its duration, but also on which age groups suffer most.

A disease's age-specific profile of morbidity and mortality is of great importance. An outbreak among children has no immediate effect on economic output, though resources will surely be diverted from other uses to treat and care for them. Output, as conventionally measured, could therefore fall in the short run, as parents stay away from work to devote more time and energy to their stricken children. There can be long-term effects on the aggregate supply of labour and skills if the epidemic is a long one and fertility does not rise to "replace" those children who die, or if some of those who survive are left mentally or physically weakened by the encounter for life.

An outbreak among young adults has wider and more profound consequences. As the illness takes hold, there is an immediate reduction in the supply of labour, but without any attendant effects on the stock of physical capital in all its forms; so that the wage rate will tend to rise unless there is a sufficiently large pool of unemployed workers. Disability or death will make reductions in the supply of labour permanent. An increase in mortality in this group is also very likely to lead to lower fertility, and hence to a smaller labour force in the future. The other channel through which an outbreak can affect the long-run supply or quality of the labour force is through its effect on the formation of children's human capital. The end result is a reduction, not in the size of the future workforce, but rather in its quality. Determining what happens to the accumulation of physical capital per worker is a complicated matter. Higher mortality

among young adults and the lower fertility that attends it will reduce the rate of growth of the labour force. At the same time, a marked and prolonged increase in morbidity will very likely divert savings away from investment in all types of capital into treatment of the sick; and the loss of lifetime family income will normally reduce savings. Hence, whether such an epidemic induces an increase or a fall in the level of physical capital per worker remains an open question.

An outbreak among the elderly has a direct effect on labour supply only to the extent that the elderly work or devote some of their time to caring for their grandchildren and their adult children's households, whereby the diversion of resources into treatment of those who fall sick also occurs. The magnitude of the long-term effects of morbidity and mortality among the old on labour supply and its quality will depend on the strength of two factors, which pull in opposite directions. There is the extent to which these contributions to younger households further fertility and the formation of human capital. At the same time, there are the claims of the old on resources, which compete with investment in human and physical capital alike.

To complete this brief account of the general workings of infectious diseases on the economy, two are particularly worthy of attention. If the scale and gravity of an outbreak are large and enduring, the whole structure of insurance arrangements could be burdened to the point of collapse. This applies not only to the capacity of the health system itself to deal with a flood of the sick and dying, but also to the very institutions that provide public services and other forms of health insurance. Affluent societies, with adequate public revenues, are better placed to deal with such a shock than poor ones, which rely heavily on the extended family and local communities to provide support to the needy in times of difficulty. The collapse of these institutions in poor countries could well result in very large economic and social costs indeed.

An outbreak of an infectious disease does not have to occur on a grand scale, however, in order to produce large economic effects. For economic activity depends heavily on expectations and psychological sentiment, the formation of which can, in turn, be very susceptible to reports of an outbreak of disease. It may take a while, depending on the nature of the disease and the location of the outbreak, before the true scale of the threat becomes clear, so that expectations in the interim can err in both directions. Should the appearance of a new disease or a new variant of an old one prove to have

grave and enduring consequences for morbidity and mortality in a population, the associated risks will eventually be embodied in the population's expectations and hence in its members' decisions concerning the allocation of resources. Erroneous or not, expectations play a central role.

3.2. Prevention, treatment and R&D

An ounce of prevention, according to the old adage, is worth a pound of cure. If true—and there is strong, presumptive evidence in its favour—it tells us where the first priority in combating all communicable diseases lies. It also suggests, on first reflection, that R&D efforts should concentrate on improving the means and methods available to prevent outbreaks of communicable diseases rather than those to treat individuals who fall sick. We now examine these general propositions in some detail.

The apparatus of prevention is made up of components that range from administration, surveillance and control to physical infrastructure, and even include, in certain cases, treatment. Organisational capacity is especially important. A public health system must undertake continuous and comprehensive monitoring and careful preparation for action, as well as possessing the resources to act when needed, if it is to function well. The ability of the surveillance apparatus to spot outbreaks and track the patterns in which they are spreading is one key element in determining the efficacy of the system. The second element is made up of diverse, complementary measures, all of which are essential in preventing infectious diseases and controlling outbreaks when they do occur: the control of disease vectors, the vigorous implementation of vaccination programmes, unrelenting public information campaigns, the imposition of isolation and quarantine, and the promotion of health education. In order to control sexually transmitted diseases, the tracking of partners is also combined with treatment, as treatment of such cases stems the spread of the disease. Even when isolation and quarantine are deemed essential, the prospect of receiving treatment while so confined will make these measures easier to enforce. At the "hard" end of the range of public interventions, the physical infrastructure that provides clean water and sanitation is vital in controlling gastro-intestinal diseases.

When does treatment become a socially profitable option in its own right, as opposed to providing merely a valuable component of a package of prevention and

control measures? No general answer can be given, for much depends on the disease in question; but the following example is important, topical and revealing.

One intervention that commends itself in connection with the control of all sexually transmitted diseases is to target prostitutes and their clients, whereby the use of condoms is also strongly promoted. Marseille, Hofmann and Kahn (2002) give the corresponding cost of averting a single case of AIDS in Kenya, for example, as US $8 to $12. This is cheap indeed, but in the nature of the disease and people, it must be inferred that this is an expenditure that will recur annually. They also present evidence that other preventive measures, such as ensuring a safe blood supply and treating mothers at birth with nevirapine, are less cost-effective by a factor of 10 or more. Choosing a bundle of diverse preventive measures, they estimate the resulting cost per disability-adjusted life-year (or DALY, see Section 3.3) so saved at $12.50.

We now move the other end of the range, where the overwhelming bulk of expenditure goes to treating those infected with the AIDS virus. Such treatment covers not only opportunistic infections, especially in the later stages of the disease, but also anti-retroviral therapies. These measures keep infected individuals healthier and can extend their lives for a few years, thereby raising lifetime family income and improving parental care. Marseille *et al*. (2002) put the cost of saving a DALY by these means at $395, on the conservative assumption that the drugs take the form of low-cost generics and that the costs of the technical and human infrastructure needed to support an effective (HAART) regime of this kind can be wholly neglected.

At this point, the reader will hanker after some idea of how much a comprehensive HAART programme might cost. The elements of an estimate for Burkina Faso, a poor West African country in which the HIV-prevalence rate is about 8%, are set out by the World Bank (2003). If generic anti-retroviral drugs can be purchased from Indian firms, the annual cost of treating each individual would be about $810; under the next best, negotiated alternative, they would more than double, to $1,730. At the prevailing prevalence rate, the lower of the two estimates, translates into an aggregate outlay that is about 80% of the Health Ministry's current budget, or about 1.8% of GDP. In Kenya, where the prevalence rate is about 15% and the level of GDP per head is similar, the aggregate outlay would be roughly twice as large. These are sobering numbers, but they are broadly in line with those emerging from the optimum programmes derived by

Bell, Devarajan and Gersbach (2003) for South Africa. Whether expenditures for this particular purpose and on this scale can be justified when there are numerous alternatives for preventing the main communicable diseases, however, seems doubtful.

We turn to priorities in the field of R&D. The general presumption that prevention should take precedence over treatment when resources are very scarce does not, in fact, necessarily imply that R&D activities should concentrate on prevention. The reason is as follows. It may be easier to develop new treatments or improve old ones than to make similar progress in developing and improving the methods and means of prevention. If the difference is sufficiently large, it will more than offset the presumed social advantage of preventive measures. Without some fairly firm knowledge of these prospective (intermediate) pay-offs to R&D, no conclusions can be drawn about the right emphasis in R&D activities with regard to prevention and treatment.

Again, an example may help to illustrate what is involved. No one would deny that an effective vaccine against HIV/AIDS would do a great deal to stem, and then reverse, the tide of this modern plague. The problem is that very serious efforts have already been underway for almost 20 years without yielding any real glimmer of success. Some expert opinion holds that the prospects of developing such a vaccine within the next 10 to 15 years are not encouraging, simply because the virus mutates so readily, and some doubt that it is possible at all. If this pessimistic assessment is indeed correct, then it might well be better to divert these efforts into developing new anti-retroviral drugs, or those designed to treat TB or malaria. Indeed, when one recalls the argument in Section 3.1 that all proposals should be subjected to a common procedure of evaluation and a common yardstick, then the development of a vaccine against malaria (another tough nut) is a natural alternative. Going one step further, one need not shy away from asking whether R&D should be pared down in favour of the free or subsidised distribution of impregnated mosquito nets, for example.

3.3. The pay-off measure

The discussion in Section 3.1 leads naturally to the proposal that one should measure the pay-off that arises from spending on combating infectious diseases as the resulting savings in “costs”, suitably defined and discounted. The task of estimating these costs is greatly complicated by the need to cover the two principal externalities, namely,

the long-term effects of morbidity and mortality on the formation of human capital and the fact that a communicable disease is a common property resource; but estimating them is still possible in principle, even though it is a daunting undertaking in practice.

When confronted with the need to choose a measure, economists frequently resort to the DALY. One DALY can be thought of as one lost year of healthy life or a certain period lived with disability, where the length of the period is adjusted for the severity of the disability. The burden of disease is then measured as the gap between current health status and an ideal situation where everyone lives into old age free of disease and disability throughout. DALYs are calculated using age-specific data on death and disability rates, and the weights attached to various disabilities are based on the valuations tendered by households in surveys designed for this purpose (WHO, 2003).

An alternative is, in effect, to weight the number of lost life-years by productivity, as measured by wage rates or earnings from labour. This can be called the expected lifetime earnings measure. Jamison, Sachs, and Wang (2001) employ this measure to assess the economic costs of the AIDS epidemic in Sub-Saharan Africa. To give a specific example for Botswana, they calculate the expected lifetime income of a 22-year-old male with 12 years of education as $\Sigma_t[S(t)Y_{12}(t)(1 + r)^{-t}]$, where $S(t)$ is the probability of surviving t years from the age of 22, $Y_{12}(t)$ is the age-earnings function with 12 years' education, and r is a constant discount rate.

In the course of the discussion in Section 3.1, it became clear that an unavoidable ethical problem arises in any procedure involving life-years, weighted or otherwise, as a unit of measure. The question can be formulated as follows: once a child is born, should the reference case be the full life-cycle into a ripe old age? If the answer is "yes", then the individual is effectively treated as irreplaceable, in the sense that any deviation from this ideal involves a permanent loss that cannot be made good. This standpoint runs into the difficulty that "replacement" fertility behaviour in connection with child mortality is a well-established feature of human demography. If, on the contrary, the answer is "no", then the act of preventing the premature death or disability of an existing individual must be compared with the contribution that a newly born member of the community would make. This problem will reappear when we discuss the importance of the age-specific mortality profiles of different diseases for the allocation of aid resources.

In addition to this general problem, both of the above pay-off measures run into their own specific difficulties. The phase of childrearing and education in the lifecycle make claims on resources before the individual emerges as a productive adult who generates resources. This fact cannot, by definition, be incorporated into the DALY framework, which normally assigns equal weights to savings of DALYs at all ages, and certainly does not allow for the use or generation of resources that are not expressed in DALYs.

This weakness is not necessarily shared by the approach based on expected lifetime earnings, which assigns, in principle, a specific weight to each age. It can be expressed on a net basis as follows: at each age, one calculates for each individual the discounted present value of future earnings saved by preventing disability and premature death, and then subtracts the discounted future costs of investment in human capital formation. The calculation of a whole set of such profiles is an arduous task and, to our knowledge, has not been carried out in full. In practice, the calculations relate to an average individual over some part of the lifecycle and are made on a gross basis, as in the study by Jamison, Sachs, and Wang (2001), described above. The actual application of the lifetime earnings approach therefore suffers similar weaknesses in measuring the pay-offs to spending on health in general and to aid for health in particular.

Neither of these approaches incorporates the positive effects of reduced adult mortality and morbidity on the formation of human capital in the future, and hence on future productivity. A rigorous analysis of these effects requires a dynamic, general equilibrium framework, an example of which is developed at length in Bell, Devarajan and Gersbach (2003). Forcing them into a DALY or related framework is not, moreover, an obviously attractive way of proceeding.

4. How to allocate aid I: prevention and treatment

Having set out the general framework in Section 3, we turn to the question of how to devise a programme of aid that will bring about a "good" allocation. We are confronted at once with the problem of fungibility; the offer of an aid programme, even if drawn up jointly as an element of an overall plan for the health sector, gives the recipient government opportunities to shift resources into other activities. Where R&D is

concerned, the scope for such "substitution" is limited, because the capacity to undertake R&D activities is concentrated overwhelmingly in developed countries. Difficulties of other kinds arise instead, so we defer a discussion of R&D until Section 5. The proposals that follow address prevention and treatment in the light of potentially severe problems of fungibility.

The fungibility of aid in general, and aid in the health sector in particular, will differ across countries, depending on the terms of donors' conditionality, the functioning of internal budgetary mechanisms, and other factors. Feyzioglu, Swaroop, and Zhu (1998) investigate the extent of aid fungibility in a panel of countries for the period 1971-90. Their findings reveal a mixed pattern. First, at the aggregate level, an increase of $1 in foreign aid leads to an increase of $0.33 in total government spending, with the remaining aid being used for tax relief.[3] Second, concessionary loans (total foreign aid less grants) in the health sector seem to be more than fully fungible at the central government level: a $1 increase in health-specific aid reduces central government expenditure on health by $1.33 (though the standard error of the estimate is large). However, the same increase in aid reduces total public investment in the health sector by only $0.35.[4] There are also interesting "cross" effects. Concessionary loans to the transport and communication sector (again, excluding grants) appear to stimulate public spending on health. Finally, the authors also assess the impact of foreign aid to the health sector on infant mortality. Concessionary loans in the health sector have a statistically significant and negative effect on infant mortality. At the mean value of such loans, a doubling thereof would reduce infant mortality by 1%. In sum, these results suggest that aid flows into the health sector are to some extent fungible, but not to such an extent that they appear to render such aid ineffectual.

We have argued from the outset that combating infectious diseases should have first priority in poor countries. If this is granted, then the next step is to draw up a plan for doing so, not just for a few years, but for the long haul. The details will naturally differ

3. This result seems to be sensitive to the sample employed. A reduced sample yields the finding that an increase of $1 in foreign aid leads to an increase of $0.95 in total government spending, suggesting a markedly lower degree of tax relief.

4. Due to the unavailability of data on donor grants at the sectoral level, Feyzioglu, Swaroop and Zhu (1998) measure foreign aid as concessionary loans. They argue that a positive correlation between grants and loans would lead to an upward bias in the estimated fungibility coefficient.

from country to country, but some international co-ordination is needed in view of the fact that the diseases and their vectors recognise no national frontiers. In drawing up such plans, we have also argued that prevention should take precedence over treatment whenever funds are sufficiently scarce. For each level of funding, the components of the corresponding plan are then subjected to cost-benefit analysis in order to ensure efficiency, and revisions are made in the light of the results.

The principal proposal is as follows: design a programme to prevent and contain the infectious diseases that inflict the most damage, subject to the condition that the programme's marginal projects yield a satisfactory rate of return (at shadow prices) in relation to other uses of aid. Provide this programme with aid, full funding and technical assistance. Then monitor and evaluate its implementation.

In all likelihood, implementing this proposal will involve far heavier spending in this sphere than any programme that is likely to emerge from the domestic political economy. The proposal is also conservative, in the sense that given the targeted goal, it implicitly rests on the assumption that each dollar of aid allocated to achieving it will induce the recipient government to withdraw $1 from this sphere for use elsewhere. If pursued in full, the proposal will, by definition, attain the desired goal. What can be said with near certainty, however, is that the entire programme will not yield a marginal social return equal to that yielded by its marginal project. When faced with such an offer, the recipient government will almost surely respond by spending more on hospitals equipped to treat cases of cancer and heart disease, or on a new airport terminal, or even less worthy items; so that the marginal project in the public sector as a whole will yield less. This is the unavoidable consequence—and cost—of fungibility.

Comprehensive funding and technical assistance cover, by definition, all direct local expenditures and imports. Given the focus of this paper, we need to examine what sort of individual components are involved in the light of the fact that aid is supposed to finance them. As noted in Section 3.2, the foundation of public health is a well-functioning apparatus of prevention, surveillance, monitoring and control. Developed countries possess such systems, and they can transfer the knowledge, expertise and techniques to build up systems elsewhere, working with the WHO as needed. Where local skills are lacking, local people can be trained, eventually in sufficient numbers. These things have been done in the past; but future efforts under the above proposal will be

on a much larger scale, and will demand a correspondingly enhanced administrative capacity. The same holds for the physical structures and equipment (laboratories and other testing facilities, clinics, vehicles, dusters and sprayers), and their complementary consumable supplies (vaccines, drugs, needles, insecticides, condoms, mosquito nets and the like).[5] To emphasise once more, all salaries must be paid as well.

These are many of the obvious items on the list. There are others that also deserve special mention. Public awareness campaigns are vital weapons in the armoury of public health measures—witness, for example, the success enjoyed by Brazil and Thailand in their efforts to contain the spread of AIDS. It follows that donors should pay for the design, production and placement of advertisements and announcements in newspapers, on radio and television, and on billboards. They should also do the same for the development and teaching of classes in public health in schools. At the other, "hard" extreme, there is the task of bringing physical infrastructure that produces potable water and sanitation to standards that correspond to the benchmark rate of return defined above.

Two final remarks are in order. First, if the total cost of implementing the above proposal is sufficiently modest, serious consideration can be given to those forms of treatment that are not closely connected with prevention and control—subject, of course, to the same evaluation procedure. The alternative is to commit more resources to R&D, a topic we take up in Section 5. Second, much is sometimes made of the losses of trained medical personnel through emigration to rich countries. To the extent that public health is a rather specialised, very small and well-furnished branch of those countries' health systems, the chances of heavy losses seem rather remote.

5. How to allocate aid II: R&D

Outputs of R&D activities have the characteristics of public goods: once invented, a new medical treatment can be provided to any patient in need at the marginal cost of producing it. In addition, the public good nature of R&D outputs largely extends

5. We do not go into the question of whether there should be user charges for individual items, though it is clearly an important consideration in drawing up any health sector "plan". The fact that significant externalities are involved rather yields a presumption in favour of subsidies.

beyond national borders, although new treatments sometimes need to be adapted to local climatic conditions or economic and social circumstances. As noted in every public economics textbook, the provision of public goods is also subject to market failure. Without public intervention, private enterprises will engage in too little or no R&D, as they cannot appropriate the entire returns from their inventive activity. There are two classic solutions to this problem: public provision of R&D and the creation of incentives for private R&D. We will discuss these two options in the context of allocating aid flows.

Consider first the use of aid money to finance public R&D. The case for aid arises from the dissimilar burden of diseases faced by rich and poor countries. As Table 2 clearly reveals, the leading causes of health problems in developing countries with high mortality rates include AIDS, malaria, perinatal conditions, respiratory infections, and diarrhoeal diseases. However, these diseases do not feature among the 10 leading health problems in developed countries. At the same time, most public health research is conducted by rich countries, and it is oriented mainly towards the health problems of correspondingly affluent patients. Aid specifically allocated to R&D into the diseases that particularly afflict poor countries can help redress the direction of research. Yet there are limits to what research in public sector institutions can achieve. R&D capacity in the North is largely concentrated in the private sector, and the financial resources necessary for engaging in the risky process of developing promising molecules into new drugs and vaccines lie outside the envelope of current aid budgets.

This brings us to the second form of public intervention, namely, the creation of incentives for private R&D, classically through the grant of invention patents. In developed country markets, patents afford pharmaceutical companies time-bound, exclusive property rights for newly invented medicines, a monopoly that allows them to charge prices above the marginal cost of manufacturing drugs and thereby to recover their outlays on R&D. In developing countries, the availability of patent protection varies, but a major change has been brought about by the Uruguay Round of Trade Negotiations (1986-94), which established new rules governing the protection of intellectual property, including the obligation for WTO members to protect pharmaceutical patents for a period of 20 years. This obligation did not apply retroactively to drugs already on the market; and given the delay in bringing promising new medicines to the market, the impact of the WTO's drug patent rules will be felt mainly after 2005. At the same time, governments

Table 2.
Leading burden of diseases in developed and developing countries, 2002

	Developed countries	**% total DALYs**
1	Ischaemic heart disease	9.1%
2	Unipolar depressive disorders	7.3%
3	Cerebrovascular disease	6.4%
4	Alcohol use disorders	3.6%
5	Hearing loss, adult onset	2.8%
6	Chronic obstructive pulmonary disease	2.6%
7	Road traffic accidents	2.5%
8	Trachea, bronchus, lung cancers	2.4%
9	Alzheimer and other dementias	2.3%
10	Self-inflicted injuries	2.3%

	Developing low mortality countries	**% total DALYs**
1	Unipolar depressive disorders	6.0%
2	Perinatal conditions	5.9%
3	Cerebrovascular disease	5.0%
4	Road traffic accidents	3.7%
5	Ischaemic heart disease	3.3%
6	Chronic obstructive pulmonary disease	3.1%
7	Lower respiratory infections	2.6%
8	Tuberculosis	2.4%
9	Diarrhoeal diseases	2.4%
10	Cataracts	2.4%

	Developing high mortality countries	**% total DALYs**
1	HIV/AIDS	9.2%
2	Lower respiratory infections	8.5%
3	Perinatal conditions	8.0%
4	Diarrhoeal diseases	5.8%
5	Malaria	5.1%
6	Maternal conditions	3.1%
7	Unipolar depressive disorders	3.1%
8	Ischaemic heart disease	2.9%
9	Measles	2.8%
10	Tuberculosis	2.7%

Source: World Health Organisation (2003).

retain the option to override exclusive patent rights by granting so-called compulsory licenses to generic drug producers. In recent trade negotiations, WTO members clarified this option and created a mechanism by which countries with insufficient domestic pharmaceutical manufacturing capacity can make use of compulsory licenses through importation. The extent to which compulsory licenses will be used as new drugs progressively take on a larger share of the market remains uncertain. It is also important to note that, for the foreseeable future, WTO rules do not require the least developed countries to provide patent protection for pharmaceutical products, although several do in their domestic laws.

From the perspective of allocating development aid, a key question is whether the funding of treatment programmes and the associated purchases of medicines, combined with the enforcement of patent rights in developing countries, would increase the effective demand for drugs, and on such a scale as to provide new incentives for private R&D. If the answer is in the affirmative, a second important question is whether new R&D activities would be devoted to diseases that mainly afflict the inhabitants of poor countries. On both questions, there are reasons to be pessimistic. Even sanguine forecasts of future aid flows to the health sector pale in comparison with the total annual revenues (over $270 billion) of research-based pharmaceutical companies (Fink, 2003). The greatest financial returns to R&D will continue to be reaped from drugs that address the health problems of patients in the North. Developed countries account for more than 80% of worldwide sales of the research-based pharmaceutical industry, with the remaining share coming mainly from upper-middle income countries. The whole of Africa contributes only 1% of the worldwide sales of this industry (Kremer, 2002). In keeping with these overwhelming differences in market size, patenting related to tropical diseases has never exceeded more than about 0.5% of overall pharmaceutical patenting (Lanjouw and Cockburn, 2000).

There are additional problems to be considered in relation to pharmaceutical patents. As already pointed out, under normal market conditions, prices for drugs under patent protection will exceed marginal costs of production. For any given aid budget, higher prices for drug treatment will reduce the number of patients benefiting from aid. If higher prices were to stimulate R&D and lead to the eventual discovery of vaccines or more effective drug treatments, this might be justified as an inter-generational trade-off. Even then, it would involve a difficult ethical choice. If, in contrast, higher prices

were to bring few future benefits to the populations of poor countries, and the resulting revenues were used, instead, to finance research into, for instance, Alzheimer's disease or erectile dysfunction, then aid flows would involve an element of subsidy to developed country-specific R&D.

It is sometimes suggested that so-called tiered or differential pricing, whereby pharmaceutical companies charge "lower" prices in poor countries, offers a way out of this dilemma. This proposition is not wholly convincing. First, if national or regional markets can be perfectly segmented, it is indeed in a profit-maximising monopolist's interest to price-discriminate according to the elasticity of local demand (although it is far from clear whether local demand elasticities fall with per capita incomes). But even in the lowest-price market, prices would still contain a mark-up over marginal cost. Second, pharmaceutical companies may decide to offer patented drugs at prices equal to marginal cost in poor countries, for example, for philanthropic reasons. While this decision would obviously resolve the pure pricing dilemma, it would not allow for any recovery of R&D costs and would therefore raise the question of why patents should be protected in the first place.

Finally, several economists have proposed strengthening the private incentives to undertake R&D activities directed at the specific diseases burdening developing countries through the provision of guaranteed purchases of new vaccines or drug treatments that meet certain standards—explicit guarantees that are to be offered by donor governments (see, for example, Kremer, 2002).[6] The rationale for such guarantees is to reduce uncertainty about future demand and thereby lower the risks of investments in these lines of R&D, thus making them relatively more attractive to other lines. This approach has the additional advantage that no public funds would be disbursed until successful products actually emerged. If and when they do, the aid money spent on the vaccines and drug treatments in question would have helped to induce successful R&D specific to developing countries, rather than general R&D or even developed country-specific R&D. Compared to public research grants, which subsidise research inputs, purchase commitments may also avoid moral hazard problems between donors and grant recipients, as they explicitly define a research output.

6. The Center for Global Development and the Global Health Policy Research Network (2004) outline a specific proposal for a purchase guarantee mechanism in the area of vaccines.

Notwithstanding these potential benefits, the effectiveness of purchase guarantees depends critically on the extent to which complete contracts can be specified and enforced. For example, while a vaccine against a particular disease may seem to be a well-defined product, few vaccines actually reach the ideal of easy inoculation followed by complete lifetime protection. Contracts would need to spell out minimum efficacy levels in well-defined target populations, the envisaged length of protection, an acceptable level of side effects, and allowed interferences that could reduce or obliterate protection. Usage characteristics, such as the route of immunisation, storage needs, and quality control mechanisms would also need to be stipulated. Since there is imperfect information on possible R&D outputs, it is difficult for donors to set these target requirements at socially optimal levels.

If complete contracting is not possible, problems of time-inconsistency may arise: once a company has sunk its R&D investment and developed a new vaccine or drug treatment, governments have an incentive to renegotiate their original commitment. Changes in disease conditions that reduce or obviate the need for a particular treatment accentuate this time-inconsistency problem. Donors would find it desirable to account for such changes *ex ante*, by including exit clauses or allowing for the possibility of renegotiating original purchase commitments. But a higher degree of flexibility in commitments is likely to increase firms' fears of being confronted with time-inconsistent behaviour. A commitment by donors to independent arbitration in case of disputes may help to reduce this fear, but is unlikely completely to eliminate it. While we are not aware of any actual experience with purchase commitments in the health sector, the history of incentive mechanisms used in the defence sector has shown that serious problems of incomplete contracting may well arise.

6. Concluding discussion

We have taken the position that aid in the sphere of health should be allocated broadly to deal with the main sources of the disease burden in poor countries, namely, communicable diseases, maternal and perinatal conditions, and malnutrition. This may seem natural and almost self evident; but there are also two particular, strong reasons for international engagement. First, communicable diseases are "common property resources", whose management requires concerted action among states. Second,

R&D capacity is heavily concentrated in the private sector of donor countries, whose populations suffer other ailments, mostly those of ageing. A further reason for giving high priority to combating certain of these diseases, notably AIDS and TB, is that they cause morbidity and mortality mainly among young adults in their prime years. Thus, not only do these diseases destroy existing human capital, but they also weaken the mechanisms through which human capital is transmitted to, and formed in, future generations. This last effect is not taken into account in conventional measures of the burden of disease.

These arguments—together with persuasive evidence that aid is rather fungible—lead us to the proposal that donors should fund the *whole* of the following long-term programme:

(i) Design a set of measures and projects to prevent and contain the infectious diseases that inflict the most damage, subject to the condition that the marginal project yields a satisfactory rate of return (at shadow prices) in relation to other uses of aid;

(ii) Support this "plan" with the necessary technical assistance; and

(iii) Monitor and evaluate its implementation, and revise the plan in the light of success and failure.

Each recipient country needs its own plan, and these individual plans will form the components of a global programme. In the nature of things, this undertaking will involve a joint effort by donors and recipients—and some troublesome problems of co-ordination at various levels.

Individual donors like their contributions to be visible, as the World Bank (2003b, Chapter 11) argues, rather oblivious to the mote in its own eye. How is this to be squared with the need to pool resources in order to finance an inherently joint undertaking? One possibility is that a consortium draws up an estimate of what can be thought of as the common costs of the plan, namely, its initial development and the drugs, vaccines and consumable supplies needed to implement it. Donors would then pay a common fraction of their total aid budgets into a central pot so as to fund these costs in full. Individual "visibility" would be achieved through the provision of technical assistance, special facilities, and funds for surveillance and the relevant payrolls at the country level.

The history of aid relationships will usually yield at least a few natural donor-recipient partnerships, from which consortia the two sides can hammer out a set of "contractual" groupings that exploit affinity and experience, while keeping relative contributions among donors more or less the same.

Quite apart from any such plan, a truly common enterprise arises from the scope for improving co-operation in the sphere of public health at the global level, an enterprise that donor countries should finance. Bell and Lewis (2004), in their analysis of the economic implications of epidemics, argue that communicable diseases pose a substantial threat to welfare in a so-called "globalised" world with highly mobile people and goods. One lesson from the experience of the SARS epidemic, for example, is that the arm of the WHO that deals with communicable diseases should be strengthened, both in its capacity and in its functions. An independent operations and evaluation department should monitor the performance of all member states and formally report their failures to adhere to established standards. Egregious and persistent cases should be referred to a central gremium, and thence, in the absence of the required corrective action, to the UN Security Council.

Thus far, we have concentrated rather on what donors should provide by way of resources. The resulting pay-off will depend on how efficiently they are put to use, and hence on the recipient's performance. In this connection, it is important to stress that the source of inadequate provision of health services is often not so much lack of funds as poor governance (World Bank, 2003b). For example, problems with counterfeit pharmaceutical products, which undermine drug treatment programmes, are rife in large parts of Africa. Health awareness campaigns are relatively cheap, but require the will and commitment of governments, even when mounting them means admitting past failures. In addition, the effectiveness of aid itself will depend on the quality of domestic governance. That does not necessarily mean that aid should only go to well-governed countries, as the neediest countries may also be the worst governed. But it does suggest that improving public health governance should be part of donors' aid efforts. To this end, we make two suggestions to complement the above proposal to strengthen the powers of the WHO.

First, the argument that communicable diseases pose a threat to the general welfare has a persuasive, if controversial corollary, namely, that the World Bank, the IMF

and other multilateral donors should make satisfactory performance in the sphere of public health—as measured by a basket of suitable indicators—an explicit element in the "conditionality" of their lending as a whole. To the extent that performance so measured depends heavily on good administration and governance, there would be a corresponding incentive to improve them. For all the doubts as to whether conditionality is really effective—some recipient governments have been rather good at getting around conditionality targets—such provisions are worth trying. Second, there is a need for a common procedure in order to impose discipline on, and consistency in, the evaluation of specific proposals. If funds are lacking for specific projects and programmes in public health, be it for immunisation against measles or the draining of swamps, the standard procedures of social cost-benefit analysis alone should determine whether to include them in the lending programme. Rich countries, in their bilateral relations as donors, should adopt the same guidelines.

REFERENCES

Bell, C., S. Devarajan and H. Gersbach (2003), "The Long-run Economic Costs of AIDS: Theory and an Application to South Africa", World Bank Policy Research Working Paper No. 3152.

Bell, C., and M. Lewis (2004), "The Economic Implications of Epidemics Old and New", *World Economics*, *5*(4): pp. 137-74.

Bonilla-Chacin, M.E. and J. Hammer (2003), "Life and Death among the Poorest", World Bank.

Center for Global Development and Global Health Policy Research Network (2004), "Making Markets for Vaccines," Center for Global Development, Washington, DC.

Fink, C. (2003), "Implementing the Doha Mandate on TRIPS and Public Health", Trade Note No. 5, Trade Department, World Bank.

Feyzioglu, F., V. Swaroop, and M. Zhu (1998), "A Panel Data Analysis of the Fungibility of Foreign Aid", *World Bank Economic Review, 12(1)*: pp. 29-58.

Gwatkin, D. and M. Guillot (2000), "The Burden of Disease among the Global Poor: The Current Situation, Future Trends, and Implications for Strategy", World Bank.

Jamison, D.T., J.D. Sachs, and J. Wang. (2001), "The Effect of the AIDS Epidemic on Economic Welfare in Sub-Saharan Africa", Commission on Macroeconomics and Health, Working Paper No. 13.

Kremer, M. (2002), "Pharmaceuticals and the Developing World", *Journal of Economic Perspectives*, *16(4)*: pp. 67-90.

Lanjouw, J.O. and I. Cockburn. (2000), "Do Patents Matter?: Empirical Evidence After GATT", NBER Working Paper No. 7495.

Marseille, E., P.B. Hoffman, and J.G. Kahn (2002), "HIV Prevention before HAART in Sub-Saharan Africa", *The Lancet*, *359*: pp. 1851-6.

Van Dillen, S. (2004). *Different Choices: Assessing Vulnerability in a South Indian Village*, Verlag für Entwicklungspolitik, Saarbrücken.

World Health Organisation (2003), "Global Burden of Disease in 2002: Data Sources, Methods and Results", WHO-GPE Discussion Paper No. 54.

World Bank (2003a), "A Technical Guide to the Contemporary Context and Procurement of HIV/AIDS Medicines and Related Supplies under Bank Funded Programs", Washington D.C., processed.

World Bank (2003b), *World Development Report 2004: Making Services Work for Poor People*, World Bank, Washington, D.C.

Comment

by
Jacky Mathonnat
Cerdi, Université d'Auvergne

Clive Bell and Carsten Fink bring a stimulating contribution to the debate on aid allocation, with the idea of improving health in developing countries. I will structure my comments around two themes: first, remarks on a selection of specific points, followed by some reflections on questions raised by the use of the directions presented by the authors.

1. Approach and justification of aid allocation for the fight against communicable diseases

When resources are rare, the question of aid allocation based on rigorously defined criteria is, of course, fundamental. The argument made by authors, for whom the struggle against communicable diseases is a priority in aid allocation, is quite illuminating. It stresses, for one thing, the necessity of protecting current and future human capital from the negative effects of communicable diseases. It also focuses on the need for proper identification and evaluation of *all* costs, via an intergenerational perspective.

The fight against communicable diseases is particularly recognised, in several studies[1], to be important for Objective 4—"Reduction of infant mortality". And it is clearly at the heart of Objective 6 —"Fighting HIV/AIDS, malaria, and other diseases".

Beyond their general argument, the authors attract our attention to several important points. I will limit myself to addressing four of them.

1. See, for example, the works of the Bellagio Group, in particular Black R., Morris S. and Bryce, J., "Where and why are 10 million children dying every year", *The Lancet,* June 2003.

They remind us in a timely manner of the importance of *fungibility*. Measuring it is difficult, but econometric analyses confirm the existence of a partial fungibility in the health sector.[2] Paradoxically, while states and donors today admit the reality of the phenomenon, it appears that they both often have difficulty at translating the implications concretely into the health financing mechanisms they put in place. But Clive Bell and Carsten Fink are right to emphasise that the risk of fungibility is reduced in terms of volume when dealing with financing of research and development (R&D), because the developing countries generally devote few resources to it.

In the fight on communicable diseases, the superiority of *prevention* over treatment is not always so clear-cut as one could sometimes think. We cannot give a systematic response to the question of whether or not R&D expenditures should concentrate on the prevention of communicable diseases or on the treatment of them. This said, the principle remains—and this is the message of the authors—that in a situation of tightly limited resources, prevention must have precedence over treatment. But does this mean that all the exterior aid should be withdrawn from interventions involved in curing communicable diseases, which would be but a matter of public non-aid financing[3], direct payments from the sick and from insurance mechanisms? It does not appear to me that this is the position of the authors, because they consider that each project included in the aid-financed project "should have a sufficient return in relation to the other uses of the aid". Let us add that in certain countries, it could turn out necessary that, despite fungibility, a fraction (very small) of the aid also finance a few operations not entirely satisfactory with regard to the criteria proposed by the authors, if, for socio-political reasons, they facilitate the adoption and the implementation of fundamental interventions.

I share Bell and Fink's analyses of the role of aid in *supporting R&D*, whether it has to do with public research financing or setting up incentives for private research. An additional problem lies in the fact that relatively few developing countries are capable of developing significant research programmes,[4] and the aims of some of the

2. See Feyzioglu, Swaroop and Zhu (1998) in Bell and Fink, as well as Devarajan S, Rajkumar P. and notably Swaroop, V., "What does aid to Africa finance?", *Working Papers,* World Bank, 2000.

3. We of course dissociate here questions of financing from production of services.

4. See Mills, A., "La science et la technologie en tant que biens publics mondiaux : s'attaquer aux maladies prioritaires des pays pauvres", *Revue d'Economie du Développement,* No.1-2, June 2002. Among others, this is one of the justifications of the WHO's Global Forum for Health Research.

programmes that do spring up could be inappropriate for the fight against communicable diseases. Lanjouw and Cokburn's 2000 study[5] reveals that 16% of the R&D expenditures in India's medical industry were aimed at tropical diseases, and that the two most important firms concentrated their works on non-communicable illnesses.

Finally, the authors remind us of two overwhelming necessities: rigorous evaluations and dissemination of the knowledge that results from it.

2. Implicating a long-term global policy

At the end of their analysis, Bell and Fink propose that donors finance the *totality*[6] of a long-term programme conceived by:

(i) Defining a set of measures and projects to prevent and contain the main infectious diseases, on condition that the yield of the minor project be satisfactory in relation to other possible uses of the aid;

(ii) Supporting this plan with the necessary technical assistance;

(iii) Monitoring and evaluating its implementation, while revising the plan in light of the successes and failures observed;

(iv) Each beneficiary country having its own plan contributing to the global programme.

My second series of remarks will touch on this on the *level of plans by country*.

The authors insist rightly on the necessity to attack these questions in the framework of a long-term programme. But should exterior aid finance the *totality* of this programme as the authors propose? It appears to me that there are *two approaches* to consider.

5. Lanjouw J. and Cockburn, I., "Do patents matter? Empirical evidence after Gatt", *NBER Working Paper* No. W7495, 2000.

6. In italics in the text.

They require strong vigilance so that the left hand does not unravel what the right hand works so hard to realise.

The first approach is to ask if the aid must finance all types of expenditure, including the salaries and other operating expenditures necessary for the implementation of the framework. I believe, like the authors, that the answer is yes. But this brings up at least three questions:

- Firstly, we know that externally financed projects take charge of all or part of the bulk of salaries, often raising very delicate problems (level of compensation, flight of skills from the administration to projects, etc.);

- Next, we must ask if the existence of projects and programmes fully financed by donors does not carry the risk, in certain countries, of working against the appropriation of policy and strategy—which many studies show is important for aid effectiveness. This brings out, notably, the recognised role played by real dialogue on policy, and conditionality based on results and governance;

- Finally, the question of knowing whether or not there is, in a programme entirely financed by external aid, a non-negligible risk of sub-programmes taking a vertical approach. This brings de-structuring effects that we can see in certain countries in the framework of the fight against HIV/AIDS or tuberculosis, for example, and this—in proportion to certain activities of the fight against communicable diseases—is naturally a matter of the Integrated Management of Childhood Illness Program (IMCI).

The *second approach* should consist of observing whether the implementation of this plan will, at country level, absorb the majority—or even the totality—of aid in the health sector.[7] This question appears to me essential for several reasons:

- Implementing the plan proposed by the authors will be undoubtedly demanding for the administrative capacities of states, even if it is entirely financed from the exterior. The success of the programme will also depend on factors that lie

7. And whether the aid allocated to the fight against communicable diseases will thus be additional, or substitutable, to other exterior and interior resources.

beyond the strict field of what goes on in the fight against communicable diseases, that is to say, elsewhere in the health system. However, non-affected sector aid could play a substantial role in the definition, financing and implementation of the institutional reforms necessary for improving the functioning of health systems, and to stimulate efficient production of services, whatever formulas are retained.

- An important part of the implementation of the programme of an aid-supported fight against communicable diseases will seek periphery care structures and de-concentrated or decentralised administrations. They often make up a particularly weak link in health systems. One example, among others: we know that local pressure tends to push personnel to favour individual clinical services, and that it is not easy to put into place effective incentives for maintaining a decent level of normalised services for populations.[8]

- As we were reminded by Filmer, Hammer and Pritchett (2002), "health policy is dependent on the anticipated effectiveness of the public sector, with existing institutional arrangements taken into consideration. If this effectiveness is weak—and it has been extremely weak in many developing countries […] adopting strategies which intensively seek out public sector capacities is of a doubtful validity […] Kerala may be able to do it, but Bihar certainly not."[9] In short, the work to be done on institutional reforms and capacity building is both immense and fundamental. In the low-income country category, more states are closer to Bihar than Kerala. Would it not, then, be suitable to reserve a part of the aid allocated for health to use, on a case-by-case basis, for institutional reforms—even if this is done to the "detriment" of complete financing for the fight against communicable diseases, because in the institutional domain, the risk of fungibility appears very weak? The same goes for the financing of social marketing—"selling" the reforms and policy to the stakeholders, an area which governments

8. Control of infectious agents, vaccinations, distribution of vitamin A… See Hammer J. and Jack, W., "The design of incentives for health care providers in developing countries: contracts, competition and cost-control" in Audibert M, Mathonnat J., and de Roodenbeke, E., *Le financement de la santé dans les pays d'Afrique et d'Asie à faible revenu,* Karthala, p. 498 Paris, 2003. It is, by the way, one of the themes of the World Bank's 2004 World Development.

9 Filmer D., Hammer J. and Pritchett, L., "Weak links in the chain II: a prescription for health policy in poor countries", *The World Bank Research Observer,* vol. 17, No. 1, 2002.

entirely neglect most of the time despite its importance. On these questions, there is undoubtedly a *balance* to be found on a case-by-case basis.

It appears to me that it would be useful to introduce the *potential role of demand and private financing* into the framework of analysis. I share the authors' point of view that the fight against communicable diseases falls largely under the logic of subsidies. But the question of aid allocation for financing the fight against communicable diseases can be posed differently, in a way in which household financing[10] (which has aspects that we know, like capacity and the will to pay) is more or less important. This is valid not only for curative care for infectious diseases, but also for certain preventative interventions (vaccination). In an environment where the total financing (interior and exterior) will be insufficient, taking the potential financing of households into account makes it possible to reduce aid devoted to this type of intervention, and to raise it for another. All things equal, this contributes to increasing the aid's efficiency.

As far as the fight against communicable diseases is concerned, aid's results also depend on the *choice and the implementation of financing instruments.* Most of the issues related to this point are not specific to health aid, but they are, however, of a crucial importance for the success or failure of a health strategy supported by exterior financing. I will limit myself here to three points:

- Bell and Fink have underlined the importance of co-ordination between donors. Developing this line of thought, the question of the schedule of aid disbursements is essential, all the more so when the execution of public spending is conditioned by an accounts plan ("cash budget")[11]. In other words, beyond the volume of available aid, the respect of the disbursement schedule linked to the financing need schedule—for programmes or strategies—is crucial. Numerous programmes and strategies have been ruined for this reason. The "common property" formula proposed by the authors is an approach that is gaining ground, and it is a good element for a response to this problem. But it must be noted that, if not desired, the "common property" constitutes a budgetary enclave, eternally managed

10. Direct payments and assurances.

11. Paradoxical situations can result. For example, introducing an approach that is very promising in terms of bottleneck budgetisation (*Marginal Budgeting for Bottlenecks)* and efforts at more efficient programming, risk, in certain countries, being wiped out by erratic disbursements.

according to exceptional procedures. This implies simultaneous improvements of capacity and budgetary management of both finance *and* health administrations in the majority of low-income countries. A broader outlook raises the question of protecting priority expenditure and, as we know, allotment offers an inadequate response to this for many reasons. Again, institutional reforms… Let us add that, as E. Berg (2003)[12] emphasises, "the passage from project aid to budgetary support"—which is logically of a "common property" orientation—"is not very likely to increase aid effectiveness in the least advanced countries" if failing institutional capacity is not developed.

-Secondly, aid has a role to play in improving the rates of budgetary execution, which are sometimes dramatically weak—not only for investment expenditures, but also for non-salary operating costs. This is often much more prejudicial.

In conclusion, an energetic and wide-ranging fight against communicable diseases as advocated by Bell and Fink most probably implies anticipating a *utilisation of aid for stimulating the demand for services*, in this sense the impact of policy will depend, in part, on the degree of response of individual decisions to official state policy. It appears to me that when these questions are addressed, governments and donors still very often react from a viewpoint that targets poverty criteria. However, the lessons learned from the Progresa programme in Mexico[13] show the risks, and suggest replacing this targeting of poverty with a precise targeting of the populations which best correspond to the specific objectives of programmes falling under the fight against communicable diseases.[14]

12. Berg E., "Augmenter l'efficacité de l'aide : une critique de quelques points de vue actuels", *Revue d'Economie du Développement,* No. 4, December 2003.

13. De Janvry, A. and Sadoulet, E., "Dépasser Bono : comment rendre l'aide plus efficace", *Revue d'Economie du Développement,* No. 4, December 2003.

14. As is done, for example, in programmes for sex workers in the framework of the fight against AIDS.

Comment

by
Agnès Soucat
Lead Health Economist, Africa Region, The World Bank,
Washington DC [1]

Introduction

In the paper "Aid and Health", Bell and Fink are reviewing the case for directing aid flows to fight communicable diseases as a matter of priority. As pointed out by the authors, this case has been made repeatedly, most often convincingly, starting with the World Development Report of 1993.

The paper identifies three main technical points that are often used to justify a focus of aid on communicable diseases. First, it is argued that the burden of diseases on the poor—both poor countries and poor groups—is mainly attributable to communicable diseases. Second, that controlling communicable diseases has an essential public good nature, given the high level of externalities produced by this control. Finally, the "public" or "common property" nature within and across international boundaries is emphasised in the paper as a key argument for establishing common health goals between poorer and richer countries.

These points appear valid on the whole and are presented in a balanced way in the paper. It is true, however, that these arguments have been over simplified in the global debate. Shortcuts have been taken with a global line reading too often as:

1. The views expressed in this paper are entirely those of the author. They do not necessarily represent those of the World Bank Group, its Executive Directors, or the countries that they represent and should not be attributed to them.

"investing in health is important" implies "a good allocation means doing something on communicable diseases" implies "more external funding for communicable diseases is the way to go for development".

But while these messages have been politically appealing and have contributed to some results on the AIDS front, there is growing evidence that the returns on this simplified story may not be as high as expected. The story may actually be more complex than initially envisioned.

1. What is communicable when we say "communicable"?

Let us take the first technical point: "communicable and related diseases represent the highest burden on poor countries and poor groups". One of the shortcuts in the global debate is the constant omission of the second part of the sentence "and related diseases". In fact, if corrected for those "other diseases", the burden of the real "communicable diseases" decreases from 73% to 52.8% in Africa and from respectively 49.3% and 44.9% to 25.5 and 22.4% in Eastern Mediterranean and South East Asia.

Using the epidemiological definition of communicable diseases, we would conclude that half of the burden of diseases in Africa and more than three-quarters in Asia and Eastern Mediterranean are related to diseases that are non-communicable. Let us not conclude that the control of communicable diseases does not remain terribly important. Yet addressing communicable diseases is not the whole story when it comes to poor groups and poor countries.

The "other diseases"—maternal and perinatal conditions, as well as nutrition deficiencies—are not amenable to strategies focusing on communicable diseases and represent a significant proportion of the burden of diseases. Regarding malnutrition, DALY numbers also do not reflect the real story, which shows, for example, that malnutrition contributes—as an underlying factor—to 50% of deaths of children. In fact an intervention such as vitamin A supplementation has been shown repeatedly to contribute to a reduction of 30 to 40% of under-five mortality. Maternal and child health interventions including nutrition interventions and childcare should also be taken into

account. In certain contexts, including Central Asian states, other non-communicable diseases have been identified as being responsible for the main burden of diseases on the poor. A recent analysis of the decline in life expectancy in Russia concludes that about half of this decline can be attributed to the consumption of alcohol and mental distress including depression (Notzon *et al.*, 1998). The slowing of population growth and labour supply that is said to be contributing to communicable diseases makes the case for control of alcohol abuse and treatment of depression in Russia. The same holds for the "epidemic" of use of heroin in the townships of Myanmar (Chelala and Beyrer, 1999).

A key argument for more aid, i.e. that diseases such as "AIDS and TB weaken the intergenerational mechanisms by which new human capital is created" and can generate long-term positive externalities, is not specific to communicable diseases either, and could well be applied to maternal ailments, traffic injuries, as well as early deaths linked to smoking and alcohol consumption or cardiovascular diseases.

2. Even more nuance is needed on the public good or common interest argument

First, the public good or high externalities argument could be applied to non-communicable diseases as well. With increasing evidence on the impact of passive smoking, the case for smoking control as a high externalities intervention is increasingly compelling; ditto for the impact of alcohol on domestic violence, child abuse and other social consequences; the case for control of water and air chemical pollutants is even stronger as a pure public good. Thus, the "infectious" or public good feature that determines the likely externalities can also be found amongst non-communicable diseases.

Second, the public good feature of "any" intervention addressing communicable diseases needs to be examined carefully. On the prevention side, the case for HIV prevention (use of condom, information) is clear-cut; but the case for immunisation less so. Externalities due to immunisation, commonly called "herd immunity", only play at relatively high levels of coverage. Thus observed immunisation levels of 30-40% in developing countries provide little beyond individual protection for the children

vaccinated. The same factors occur for the use of *Insecticide Treated Bed Nets*: to provide benefits beyond the individual sleeping under it, a minimum level of coverage in the community—estimated to be around 40-50%—is needed.

On the treatment side, some clearly provide high levels of externalities: the treatment of syphilis or TB interrupts the transmission of the disease, and the benefits extend largely beyond the individual treated. Yet the same cannot be said of AIDS treatment, including the most recent *Highly Active Antiretroviral Therapy* (HAART). This treatment benefits the individual but only reduces transmission—reducing the viral load—without interrupting it. Increased survival, on the other hand, increases mechanically the risk of transmission. Some externalities could be nonetheless expected if the availability of treatment triggered safer behaviour; yet the current literature seems to point rather towards the occurrence of "moral hazard" associated with treatment—including recurrence of unsafe behaviour in key groups in Europe and the USA. So the economic argument for AIDS treatment relies less on the public good argument than on the economic returns of saving a productive life. This argument, however, would hold as much for alcohol and smoking prevention, as well as maternal deaths prevention and cancer treatment.

Third, the premise of a common interest of poor and rich countries alike in addressing communicable diseases can be questioned. Funding for an AIDS vaccine efficient on the particular HIV subtype found in Africa is very limited. Most of research focuses on the particular subtypes found in Northern America and Europe; in the same way funds allocated to further developing promising malaria vaccines are largely insufficient.

A particular case in point relates to the recent efforts at eradicating poliomyelitis. While a serious debilitating disease, polio is not a major killer in low-income countries—it kills much less than measles—and polio contributes little to the burden of diseases. Prior to the eradication efforts, polio had not been seen by countries as a disease of such a level of priority that it would call for significantly more investments than other child killers. Here, the global public good argument has been used to lead poor countries to invest in an activity that had not been identified as a priority by these very countries. Over the last five years, most low-income countries received levels of external aid for polio that exceeded by several fold—more than five-fold in India—the amounts they were able to allocate to other immunisation efforts, including measles.

Polio eradication will clearly have benefits for rich and middle income countries which will be able to cut down their polio immunisation cost; yet there could be legitimate questions as to the value of those benefits to poor countries, which indeed have more pressing priorities: would not eradication of measles or a focus on bed nets have produced higher levels of benefits for a similar cost? Are the rather small benefits for these countries worth the extra effort to reach eradication rather than control—control being the strategy successfully used by developed countries over the last 50 years? What is the opportunity cost of mobilising health systems—mobilising poorly staffed MOH for several months every year—around the one goal of reducing polio to a few cases, rather than other priorities? One also needs to know that there is still a relatively high level of uncertainty on the likelihood of ever reaching the eradication goal with recent outbreaks of polio in of Africa and Asia, five years after the original set date for eradication.

3. Can delivery systems really be built around the "C" word?

So what about focusing on specific communicable diseases as a principle for aid? Identifying the main causes of child mortality and designing programmes that are likely to be effective is the clear pathway for reducing child mortality, for example. Yet the rationale for focusing on the "C" word only—while ignoring nutrition and neonatal interventions is not as clear cut—. The potential of investing in interventions tackling non-communicable problems, such as micronutrient supplementation (particularly with vitamin A), or exclusive and prolonged breastfeeding, but also assistance at delivery and family planning, etc. should not be dismissed. Those also have equivalent high potential to reduce under-five and maternal mortality, for example, as tackling malaria or HIV.

Moreover, the foundation of public health service delivery is not particular to communicable diseases. Potential economies of scope in health services are large. Indeed, typical health services are almost always bundled (priority and non-priority, communicable and non communicable), as they respond to symptoms and complaints not to epidemiology; individuals complain about cough and fever, they do not knock at the door of a health clinic asking for TB treatment.

Hence, in the Benin Republic, in India or in Mexico, the same clinic or outreach session delivers both immunisation, micronutrient supplementation and antenatal care.

To be able to identify and treat patients with tuberculosis, health services need to be able to welcome all cases of cough and fever; to be able to treat HIV, services need to screen adult patients presenting various types of symptoms; to treat malaria, all cases of fever need to be considered. Public awareness campaigns are also effective when focusing on non-communicable interventions; promotion of breastfeeding in Brazil has been demonstrated to significantly curb under-five mortality (Victora *et al.*, 1989).

The delivery of low priority services may in certain cases also be the price to pay (a transaction cost) for delivering high impact services. An example is using the entry point of a simple antenatal visit—for which there is little evidence of impact—to deliver a TT vaccination or conduct a syphilis or HIV test; or to distribute a bed net.

In summary, while the economic arguments for focusing on communicable diseases have solid grounds, simplified policy messages focusing on the "C" word may be too narrow, as they can contribute to missing an important part of the epidemiological agenda, as well as the reality of service delivery on the ground. The simultaneous use of burden of diseases, public good and countries common goals criteria as a rationale for aid does not always lead to a concurrent and consensual justification of the funding of any intervention on communicable diseases. As seen above, the various rationales may be contradictory. Nor does the use of this same rationale a priori exclude interventions on non-communicable diseases. Depending on the context, a similar strong case can be made for Vitamin A supplementation, promotion of assisted delivery, smoking and alcohol addiction prevention, as for many communicable diseases.

Finally, too narrow a focus on specific AIDS and malaria activities runs the risk of ignoring the very integrated nature of most health services and the need to build pipelines that will deliver high impact services. Difficulties encountered by low-income countries in disbursing the large amounts made available for AIDS over the last three years through the World Bank and GFTAM seem to confirm that pushing the "C " agenda alone may not work if the service delivery pipes are clogged. The 2003 PRSP report of the Government of Rwanda noted that the disconnect between large amounts of funds made available for AIDS and the lack of funding of health system was starting to pose significant problems to the country, jeopardising its capacity to deliver on both AIDS and other health interventions.

REFERENCES

Notzon, F.C, Komarov, Y.M, Ermakov, S.P, Sempos, C.T, Marks, J.S, Sempos, E.V, (1998), "Causes of Declining Life Expectancy in Russia", *AMA.*, pp. 279:793-800.

Chelala, C. and Beyrer, C. (1999), *Drug use and HIV/AIDS in Burma*, Lancet, 354(9184):1119.

Victora, C.G, Smith, P.G, Vaughan, J.P, Nobre, L.C, Lombardi, C, Teixeira, A.M, Fuchs, S.M, Moreira, L.B, Gigante, L.P, Barros, F.C. (1987), "Evidence for protection by breast-feeding against infant deaths from infectious diseases in Brazil", Lancet, 2 (8554): pp. 319-22.

Evaluating the Impact of Development Aid Programmes: The Role of Randomised Evaluations

by
Esther Duflo[1]
Massachusetts Institute of Technology
Department of Economics and Poverty Action Lab

1. Introduction

The benefits of evaluating the impact of development programmes, and of knowing which programmes work, extend far beyond any programme or agency. Credible impact evaluations are global public goods in the sense that they can offer reliable guidance to international organisations, governments, donors, and NGOs beyond national borders. The lack of credible evaluation of development programmes is also often read by the public as a sign of the failure of the programmes themselves, so that impact evaluation can help build support for development.

Traditional methods of measuring programme impact may be subject to serious bias due to omitted variables. For a broad class of development programmes, randomised evaluations can be used to address these problems. Of course, not all programmes can be evaluated with randomised evaluations; for example, examinations of issues such as central bank independence must rely on other methods of evaluation. Programmes

1. I thank Abhijit Banerjee, Angus Deaton, Michael Kremer, Edward Miguel and Rachel Glennerster for helpful discussions. The paper benefited from comments by Paul Shultz and Francois Bourguignon on a related paper presented at the ABCDE in Bangalore ("Scaling up and Evaluation"), and from work on the paper "Use of Randomisation for the Evaluation of Development Effectiveness", written in collaboration with Michael Kremer and presented at the OED conference in Washington. I gratefully acknowledge financial support from the Alfred P. Sloan Foundation and the National Institute of Health.

targeted to individuals or local communities (such as sanitation, local government reforms, education, and health) are likely to be strong candidates for randomised evaluations. This paper uses educational programmes in developing countries as case study for randomised evaluations.

Historically, prospective random evaluations of development programmes have constituted a tiny fraction of all development evaluations. In this paper, we argue that there is scope for considerably expanding their use, although they must necessarily remain a small fraction of all evaluations.

We do not propose that all projects be subject to randomised evaluations. But we argue that there is currently a tremendous imbalance in evaluation methodologies, and that increasing the share of projects subject to randomised evaluation from near-zero to even a small fraction could have a significant impact on increasing our knowledge about what works in development. All too often, development policy is based on trends; randomised evaluations could allow it to be based on evidence.

The paper proceeds as follows: Section 1 discusses the methodology of randomised evaluations: we present the impact evaluation problem, review why other evaluation methods may be unable to adequately control for selection bias, and discuss why randomised evaluations can be useful in addressing the problems encountered by other evaluation practices. Section 2 reviews recent randomised evaluations in developing countries. Section 3 extracts lessons from the evaluations described in Section 2, and Section 4 reviews an example of current practice, offers political economy explanations for the failure to adopt randomised evaluations, and discusses the role the international agencies play in promoting and financing rigorous evaluations. Section 5 concludes.

2. The methodology of randomised evaluation

2.1. The evaluation problem

Any impact evaluation attempts to answer an essentially counterfactual question: how would individuals who benefited from the programme have fared in its absence?

Inversely, how would those who did not benefit have fared if they had been exposed to the programme? The difficulty with these questions is immediately apparent: at a given point in time, an individual is observed either exposed to the programme, or not. Comparing the same individual over time will not, in most cases, give us a reliable estimate of the programme's impact on him, since many other factors may have changed simultaneously with the introduction of the programme. It is not, therefore, reasonable to seek to obtain an estimate of the impact of the programme on each individual. All we can hope is to be able to obtain the average impact of the programme on a group of individuals, by comparing them to a similar group that was not exposed to the programme.

The critical objective of impact evaluation is therefore to establish a credible *comparison group*, a group of individuals who, *in the absence of the programme*, would have had outcomes similar to those who were exposed to the programme. This group gives us an idea of what would have happened to the programme group if it had not been exposed, and thus allows us to obtain an estimate of the average impact on the group in question. Generally, individuals who were subjected to the programme and those who were not are very different: programmes target specific areas (e.g. poorer or richer areas), individuals are screened for participation (e.g. using criteria such as poverty or motivation), and finally the decision to participate is often voluntary. For all of these reasons, those who were not exposed to a programme are often not a good comparison group for those who were. Any difference between them could be attributed to two factors: pre-existing differences (the so called "selection bias"), and the impact of the programme. Since we have no reliable way of estimating the size of the selection bias, we cannot decompose the overall difference into a treatment effect and a bias term.

To resolve this, programme evaluations typically need to be carefully planned in advance in order to determine which group forms a likely control group. One situation where the selection bias disappears is when the treatment and the comparison groups are selected randomly from a potential population of beneficiaries (individuals, communities, schools or classrooms). In this case, on average, we can be assured that those who are exposed to the programmes are no different than those who are not, and that a statistically significant difference in the two groups in terms of the programme's targeted outcomes can be confidently attributed to it.

As we will see later, the random selection of treatment and comparison groups can occur in several circumstances. Using the example of Progresa, a programme designed to increase school participation in Mexico, we discuss how prospective randomised evaluations can be used and how their results can help in scaling successful programmes. Then, drawing on examples of school-based health programmes in Kenya and India, we illustrate how prospective randomised evaluations can be used when implementing adapted replications of the programme. Finally, using the example of the set-asides for women in politics in India, we discuss how programme-induced randomisation can be used.

It is worth briefly making a few clarifications regarding the use of randomised evaluations to estimate programme effects. First, a point should be made about what the evaluation is attempting to estimate. Randomised evaluations can be used to estimate the effect of a treatment on either the entire population subjected to the randomisation or on a subset of the population defined by predetermined characteristics. In contrast, instrumental variable techniques estimate local average treatment effects. Second, randomised evaluations estimate partial equilibrium treatment effects, which may differ from general equilibrium treatment effects. It is possible that, if some educational programmes were implemented on a large scale, the programmes could affect the functioning of the school system and thus have a different impact.

2.2. Alternative techniques to control for selection and omitted variable bias

Natural or organised randomised evaluations are not the only methodologies that can be used to obtain credible impact evaluations of programme impact. Researchers have developed alternative techniques to control for bias, and progress has been made, most notably by labour economists.[2] Below, we briefly review some of the most pervasive techniques: propensity score matching, difference-in-difference estimates, and regression discontinuity design.

One strategy to control for bias is to attempt to find a control group that is as comparable as possible to the treatment group, at least along observable dimensions.

2. There are numerous excellent technical and non-technical surveys of these techniques, as well as their value and limitations. See Angrist and Krueger (1999 and 2001), Card (1999) and Meyer (1995).

This can be done by collecting as many covariates as possible and then adjusting the computed differences through a regression, or by "matching" the programme and the comparison group by forming a comparison group that is as similar as possible. One possibility is to predict the probability that a given individual is in the comparison or the treatment group on the basis of all available observable characteristics, and to then form a comparison group by selecting people who have the same probability of being treated as those who were actually treated ("propensity score matching"). The challenge with this method, as with regression controls, is that it hinges on having identified all potentially relevant differences between the treatment and control groups. In cases where the treatment is assigned on the basis of a variable that is not observed by the researcher (demand for the service, for example), this technique can lead to misleading inferences.

A second strategy is what is often called the "difference-in-difference" technique. When it can be convincingly argued that the outcome would not have had differentials trends in regions that received the programme, if the programme had not been put in place, it is possible to compare the *growth* in the targeted variables between programme and non-programme regions. However, it is important not to take this assumption for granted. This identification assumption cannot be tested and, even to ascertain its plausibility, one needs to have long time-series of data prior to implementation in order to be able to compare trends over sufficiently long periods. One also needs to make sure that no other programme was implemented at the same time, which is often not the case. Finally, when drawing inferences one must take into account that regions are often affected by time-persistent shocks that may look like "programme effects". Bertrand, Duflo, and Mullainathan (2002) found that difference-in-difference estimations (as commonly performed) can severely bias standard errors: the researchers randomly generated placebo laws and found that, with about 20 years of data, difference-in-difference estimates found an "effect" significant at the 5% level of up to 45% of the placebo laws.

Providing an example of the effective use of difference-in-difference estimates, Duflo (2001) took advantage of a rapid school expansion programme that occurred in Indonesia in the 1970s to estimate the impact of building schools on schooling and subsequent wages. Identification was made possible by the fact that the allocation rule for the schools was known (more schools were built in places with low initial enrolment

rates), and by the fact that the cohorts participating in the programme are easily identified (children of 12 years or older when the programme started did not participate in the programme). The increased growth of education across cohorts in regions that received more schools suggests that access to schools contributed to increased education. The trends were quite parallel before the programme and shifted clearly for the first cohort that was exposed to the programme, thus reinforcing confidence in the identification assumption. However, this identification strategy is not usually valid; often when policy changes are used to identify the effect of a particular policy, the policy change is itself endogenous to the outcomes it was meant to affect, thus making identification impossible (Besley and Case, 2000).

Finally, a third strategy, called "regression discontinuity design" (Campbell, 1969), takes advantage of the fact that programme rules sometimes generate discontinuities that can be used to identify the effect of the programme by comparing those above a certain threshold to those just below it. If resources are allocated on the basis of a certain number of points, it is possible to compare those just above to those just below the threshold. Angrist and Lavy (1999) use this technique to evaluate the impact of class size in Israel, where a second teacher is allocated every time the class size grows above 40. This policy generates discontinuities in class size when the enrolment in a grade grows from 40 to 41 (as class size changes from one class of 40 students to one class each of 20 and 21 students). Angrist and Lavy compared test scores in classes just above and just below this threshold, and found that those just above the threshold have significantly higher test scores than those just below. This can confidently be attributed to the class size, since it is very unlikely that schools on both sides of the threshold have any other systematic differences.[3] Such discontinuities in programme rules, when enforced, are thus sources of identification.

In developing countries, however, it is often likely to be the case that rules are not enforced sufficiently strictly to generate discontinuities that can be used for identification purposes. For example, researchers attempted to use as a source of identification the discontinuity in the policy of the Grameen bank (the flagship micro-credit organisation

3. Angrist and Lavy note that parents who discover they received a bad draw in the "enrolment lottery" (e.g., an enrolment of 38) might then move their children out of the public school system and into private schools. However, as Angrist and Lavy discuss, private elementary schooling is rare in Israel outside of the ultra-orthodox community.

in Bangladesh), which is to lend only to people who own less than one acre of land (Pitt and Khandker, 1998). It turns out that in practice, the Grameen bank lends to many people who own more than one acre of land, and that there is no discontinuity in the probability of borrowing at the threshold (Morduch, 1998).

These three techniques are subject to large biases that can lead to either overestimation or underestimation of programme impact. LaLonde (1986) found that many of the econometric procedures and comparison groups used in programme evaluations did not yield accurate or precise estimates, and that such econometric estimates often differ significantly from experimental results.

Identification issues with non-randomised evaluation methods must be tackled with extreme care because they are less transparent and more subject to divergence of opinion. Moreover, the differences between good and bad non-randomised evaluations are difficult to communicate, especially to policymakers, because of all the caveats that must accompany the results. In practice, these caveats might either never be provided to policymakers or simply ignored when they are; in either case, policymakers are likely to be radically misled. This suggests that while non-randomised evaluations will continue to be needed, there should be a commitment to conduct randomised evaluations where possible.

3. Examples of prospective randomised evaluations

3.1. Pilot projects

Prior to a programme being launched on a large scale, a pilot project, necessarily limited in scope, is often implemented. In most circumstances, it is possible to randomly select the beneficiaries of the pilot, since many potential sites (or individuals) are as deserving as the next. The pilot can then be used, not only if the programme proves feasible (which is the current aim of most pilots), but also when the programme has the expected impact. Job training and income maintenance programmes are prominent examples of randomised evaluations. A growing number of such pilot projects are evaluated, often through a collaboration of NGOs and academics (see, for example, Kremer 2003 for several references).

To illustrate briefly how these studies can work in practice, we analyse an example from India, taken from Banerjee and others (2001). This study evaluated an Indian NGO's (Seva Mandir) decision to hire a second teacher in their non-formal village education centres. Non-formal schools seek to provide basic numeracy and literacy skills to children who do not attend formal education. In addition, in the medium term, they aim to "mainstream" these children into the regular school system. These centres are plagued by high teacher and child absenteeism. A second teacher (often a woman) was randomly assigned to 21 out of 42 schools. The aim was to increase the number of days the school was open, to increase children's participation, and to increase performance by providing more individualised attention to the children. By providing a female teacher, the NGO also hoped to make school more attractive for girls. Teacher and child attendance were regularly monitored in programme and comparison schools during the entire duration of the project. The impact of the programme on learning was measured by testing children at the end of the school year. The programme reduced the number of days a school was closed: one-teacher schools are closed 39% of the time, whereas two-teacher schools are closed 24% of the time. Girl's attendance increased by 50%. However, there was no difference in test scores.

Carefully evaluated pilot projects form a sound basis for the decision to scale up projects. In the example just discussed, the two-teacher programme was *not* implemented on a full scale by the NGO, on the grounds that the benefits were not sufficient to outweigh the cost. The savings were used to expand other programmes. Positive results, on the other hand, can help build a consensus for projects, which hold the potential of being extended far beyond the initially envisioned scale. The Progresa programme in Mexico is the most striking example of this phenomenon. Progresa offers grants, distributed to women, conditional on children's school attendance and preventative health measures (nutrition supplementation, healthcare visits, and participation in health education programmes). In 1998, when the programme was launched, officials in the Mexican government decided to take advantage of the fact that budgetary constraints made it impossible to reach the 50,000 potential beneficiary communities all at once, and began with a pilot programme in 506 communities. Half of those were randomly selected to receive the programme, and baseline and subsequent data were collected in the remaining communities (Gertler and Boyce, 2001). Part of the rationale for launching this pilot programme was to increase the probability that the programme would be continued in case of a change in the party in power. The

proponents of the programme understood that to be scaled up successfully, the programme would require ongoing political support. The task of evaluating the programme was given to academic researchers, through the International Food Policy Research Institute. The data was made accessible to many different people, and a number of papers have been written on its impact (most of them are accessible on the IFPRI website).

The evaluations showed that the programme was effective in improving health and education. Comparing Progresa beneficiaries and non-beneficiaries, Gertler and Boyce (2001) show that children had about a 23% reduction in the incidence of illness, a 1-4% increase in height, and an 18% reduction in anaemia. Adults experienced a reduction of 19% in the number of days lost due to illness. Shultz (2001) finds an average increase in enrolment of 3.4% for all students in Grades 1 through 8; the increase was largest among girls who had completed Grade 6, at 14.8%. In part because the programme had been shown to be successful, it was continued when the Mexican government changed. By 2000, it was reaching 2.6 million families, or 10% of the entire population of families, and had a budget of US$800 million, or 0.2% of GDP (Gertler and Boyce, 2001). It was subsequently expanded to urban communities and, with support from the World Bank, similar programmes are being implemented in several neighbouring Latin American countries. Mexican officials transformed a budgetary constraint into an opportunity, and made evaluation the cornerstone of subsequent scaling up. They were rewarded by the expansion of the programme, and by the tremendous visibility that it acquired.

3.2. Expansion of existing projects

It is sometimes possible to evaluate the impact of programmes that have already shown their large-scale potential. In contrast to pilot projects, one can then ensure that the programme can be implemented on a large scale. It also may make it easier to evaluate the programme in several sites at the same time, and thus alleviate some of the concerns about external validity. A natural occasion for such evaluation is when the programme is ready to expand, and the expansion can be phased in in random order. The evaluation of a remedial education programme by Banerjee, Cole, Duflo and Linden (2003) is an example of this approach. The programme, run by Pratham, an Indian NGO, was implemented in 1994. A decade later, Pratham reached over 161,000

children in 20 cities. The programme hires a young woman from the local community to provide remedial education in government schools to children who have reached Grade 2, 3 or 4 without having mastered the basic Grade 1 competencies. Children who are identified as lagging behind are extracted from the regular classroom for two hours a day to receive this instruction.

As Pratham was seeking to expand its flagship intervention, it wanted to evaluate the impact of the programme. The expansion into a new city, Vadodara, provided an opportunity to conduct a randomised evaluation. In the first year (1999-2000), the programme was introduced in 49 (randomly selected) of the 123 Vadodara government schools. In 2000-2001, the programme was introduced to all schools, with the exception that half the schools were given a remedial teacher for Grade 3, and half given one for Grade 4. Grade 3 students in schools that received the programme in Grade 4 serve as the comparison group for Grade 3 students in schools that got the programme in Grade 3. At the same time, a similar intervention was conducted in a district of Mumbai, where half the schools received the remedial teachers in Grade 2, and half received them in Grade 3. The programme was continued for one additional year, with the school switching groups. The programme was thus conducted in several grades, in two cities, and without any school feeling that it was deprived of resources. After two years, the programme increased the average test score by 0.39 standard deviations, (which represents an increase of 3.2 points out of a possible 100; the mean in the control group was 32.4 points. It had an even stronger impact on the test scores of the children who had low scores initially (an increase of 3.7 points, or 0.6 standard deviation, on a basis of 10.8 points). The impact of the programme increased over time, but it was very similar across cities and child gender. Hiring remedial education teachers from the community appears to be 10 times more cost effective than hiring new teachers. One could be relatively confident on the basis of these estimates in recommending that the programme be scaled up, at least in India. The programme was continued for a period of time, it was evaluated in two very different contexts, and it demonstrated its ability to be rolled out on a large scale.

3.3. Programme-induced randomisation

In some instances, fairness or transparency considerations make randomisation the best way to choose the recipients of a programme. Such programmes are natural

candidates for evaluation, since the evaluation exercise does not require any modification of the design.

Allocation to particular schools is often carried out by lottery if schools are oversubscribed. In some school systems in the US, students have the option of applying to "magnet schools" or schools with special programmes, and admission is often granted by lottery. Cullen, Jacob and Levitt (2002) use this feature to evaluate the impact of school choice in the Chicago school system by comparing lottery winners and losers. Since each school runs its own lottery, their paper is in effect taking advantage of 1,000 different lotteries! They find that lottery winners are less likely to attend their neighbourhood schools than lottery losers, but more likely to remain in the Chicago school system. However, their subsequent performance is actually *worse* than that of lottery losers. This is in sharp contrast to what would have been expected and what a "naïve" comparison would have found: the results of children who attended a school of their choice are indeed better than that of those who do not. Voucher programmes constitute another example of programmes which often feature lotteries. The government allocates only a limited budget to the programme, the programme is oversubscribed, and a lottery is used to pick the beneficiaries. Angrist and others (2002) evaluated a Colombian programme in which vouchers for private schools were allocated by lottery because of the programme's budget limitations. Vouchers were renewable conditional on satisfactory academic performance. They compare lottery winners and losers. Lottery winners were 15-20% more likely to attend private school, 10% more likely to complete Grade 8, and scored 0.2 standard deviations higher on standardised tests, equivalent to a full grade level. Winners were substantially more likely to graduate from high school and scored higher on high school completion/college entrance exams. The benefits of this programme to participants clearly exceeded the cost, which was similar to the cost of providing a public school place.

When nationwide policies include a randomisation aspect, this provides a unique opportunity to evaluate a policy that has already been scaled up in several locations. The knowledge gained from this experience can be used to inform policy decisions to expand the policy in the countries, to continue with the programme, or to expand in other countries. However, because the randomisation is built into the programme design, rather than being included as an evaluation measure, the data on which the evaluation is based are not always available. International agencies can play two key

roles in this respect: first, they can organise and finance limited data collection efforts; second, they can encourage governments and statistical offices to link existing data sources that can be used to evaluate the experiments.

Set-asides for women and minorities in the decentralised government (the Panchayat system) in India are an interesting example. In 1993, the 73rd Amendment to the Constitution of India required the States to set up a three-tiered Panchayat system (village, block, and district levels), directly elected by the people, for the administration of local public goods. Elections must take place every five years and Panchayat councils have the latitude to decide how to allocate local infrastructure funds. The amendment also required that one-third of all positions (of council members and council chairpersons) be reserved for women and that disadvantaged minorities (certain castes and tribes) be proportionately represented. To avoid any possible manipulation, the law stipulated that the reserved positions would be randomly allocated. Chattopadhyay and Duflo (2001) evaluated the impact in West Bengal of the reservation of seats for women. They collected data on 465 villages in 165 councils in one district, and they found that women tend to allocate more resources to drinking water and roads and less for education. This corresponds to the priorities expressed by men and women through their complaints to the Panchayat authorities. Before completing a second draft of this paper (Chattopadhyay and Duflo 2003), they collected the same data in a poor district of Rajasthan, Udaipur. They found that there, women invest more in drinking water and less on roads, and that this corresponds again to the ordering of complaints expressed by men and women. These results were obtained in two very different districts with different histories (West Bengal had a Panchayat since 1978, while Rajasthan had none until 1995; Rajasthan is also has particularly low female literacy), suggesting that the gender of the policymakers matters both in more and less developed political systems. Furthermore, it provides indirect (but powerful) evidence that local elected officials *do* have power, even in relatively "young" systems. They also evaluated the impact of reserving seats for certain castes, and found that a larger share of goods gets attributed to their hamlets when the head of a Panchayat is from the targeted caste.

In principle, data exist that would permit us to evaluate the impact of this experiment on a much larger scale. Village-level census data is available for 1991 and will soon become available for 2001. The National Sample Survey Organisation (NSSO) conducts

large-scale detailed consumption and labour surveys every five years, with detailed data on outcomes. However, administrative barriers make this data very difficult to use: the census does not indicate to which village each Panchayat belongs. The information about Panchayat reservation and composition is not centralised, even at the State level (it is available only at the district level). Likewise, the NSS contains no information about the Panchayat. This is an example where, at a relatively small cost, available information could be made available to evaluate a very large programme. It requires co-ordination of various people and various agencies, a task that the international organisations should be well placed to accomplish.

4. Lessons

The evaluations described in Section 2 offer both substantive and methodological lessons. Some of the latter that can be drawn from the examples discussed in Section 2, are discussed below.

4.1. Results from randomised evaluations can be quite different from those drawn from retrospective evaluations

When the evaluation is not planned *ex ante*, in order to evaluate the impact of a programme, researchers must resort to before and after comparisons (when a baseline was conducted), or comparisons between beneficiaries and communities that, for some reason, were not exposed to the programme. When the determinants of exposure to the programme are not known (or worse, when they are known to be likely to introduce selection bias), those comparisons are likely to be biased. The data collection is often as expansive as for a randomised evaluation, but the inferences are biased. As we have argued above, controlling for observable differences between treatment and control groups (through a regression analysis or through propensity score matching) will correct for the bias only if it is known with certainty that beneficiaries and non-beneficiaries are comparable in these characteristics. This is unlikely to be true unless the programme was randomly allocated conditional on these characteristics. In particular, a project officer trying to optimally allocate a programme typically has more information than a researcher, and will (and should) make use of it when allocating the resources.

These concerns have serious practical implications. Studies comparing experimental and non-experimental estimates with the same data show that the results from randomised evaluations can be quite different from those drawn from non-randomised evaluations. In a celebrated analysis of job training programmes, LaLonde (1986) found that many of the econometric procedures and comparison groups used in programme evaluations did not yield accurate or precise estimates, and that such econometric estimates often differed significantly from experimental results. Comparative studies that estimate a programme's impact using experimental methods and then re-estimate impact using one or several non-experimental methods suggest that omitted-variable bias is a significant problem beyond just the examples mentioned here. Although we are not aware of any systematic review of studies conducted in developing countries, one recent study in developed countries suggests that omitted-variable bias is a major problem when non-experimental methods are used (Glazerman *et al.*, 2002). This study assessed both experimental and non-experimental methods in the context of welfare, job training, and employment service programmes and found that non-experimental estimators often produce results dramatically different from those of randomised evaluations. Further, it found that the estimated bias is often large, and that no strategy seems to perform consistently well.[4]

We are not aware of a systematic meta-analysis for developing countries, but there are many examples in which it is possible to compare randomised evaluations with prospective evaluations using the same dataset. Glewwe and others (2003) compared retrospective and prospective analyses of the effect of flip charts on test scores. Retrospective estimates using straightforward OLS regressions suggest that flip charts raise test scores by up to 20% of a standard deviation, robust to the inclusion of control variables. On the other hand, difference-in-difference estimates suggest a smaller effect of about 5% of a standard deviation. Though this effect is still significant, sometimes this is only at the 10% level. In contrast, prospective estimates based on randomised evaluations provide no evidence that flip charts increase test scores. These results suggest that using retrospective data to compare test scores seriously overestimates the charts' effectiveness. A difference-in-difference approach reduced

4. One recent study not included in Glazerman, Levy, and Meyers (2002) is that of Buddlemeyer and Skoufias (2003). Buddlemeyer and Skoufias use randomised evaluation results as a benchmark to examine the performance of regression discontinuity design for evaluating the impact of the Progresa programme on child health and school attendance and find the performance of regression discontinuity design in this case to be good.

but did not eliminate the problem and, moreover, it is not clear that such a difference-in-difference approach has general applicability. These examples suggest that OLS estimates are biased upward, rather than downward. This is plausible, since in a poor country with a substantial local role in education, inputs are likely to be correlated with favourable unobserved community characteristics. If the direction of omitted variable bias were similar in other retrospective analyses of educational inputs in developing countries, the effects of inputs may be even more modest than retrospective studies suggest. Some of the results are more encouraging: for example, Buddlemeyer and Skoufias (2003) used randomised evaluation results as a benchmark to examine the performance of regression discontinuity design for evaluating the impact of the Progresa programme on child health and school attendance. The researchers found the performance of regression discontinuity design in this case to be remarkably good: impact estimates with this quasi-experimental method agreed with experimental evidence in 10 out of 12 cases, and the two exceptions both occurred in the first year of the programme. Such research can provide invaluable guidance regarding the validity and potential biases of quasi-experimental estimators.

Future research along these lines would be valuable, since such comparative studies can help show the extent to which the biases of retrospective estimates are significant. However, when the comparison group for the non-experimental portions of these comparative studies is decided *ex post*, the evaluator may be able to pick from a variety of plausible comparison groups, some of which may have results that match experimental estimates and some of which may not (as discussed below, this is also an issue for retrospective studies in regard to problems with publication bias). Possible ways of addressing these concerns in the future include first conducting non-experimental evaluations, before the results of randomised evaluations are released, or having researchers conduct blind non-experimental evaluations without knowledge of the results of randomised evaluations or other non-experimental studies.

4.2. Randomised evaluations are often feasible

As we noted in the introduction, randomised evaluations are not suitable for all types of programmes. They are suitable for programmes that are targeted to individuals or communities, and where the objectives are well defined. For example, the efficacy of foreign aid disbursed as general budget support cannot be evaluated in this way. It may

be desirable, for efficiency or political reasons, to disburse some fraction of aid in this form, although it would be extremely costly to distribute all the foreign aid in the form of general budget support, precisely because it leaves no place for rigorous evaluation of projects. However, as the examples in this paper demonstrate, in many cases, randomised evaluations are feasible. The main cost is that of data collection, and this is no larger than that of collecting any other data. In fact, by imposing some discipline on which data to collect (the outcomes of interest are defined *ex ante* and do not evolve as the programme fails to affect them) may reduce the cost of data collection.

Political economy concerns sometimes make it difficult to not expand a programme to the entire population, especially when its success has already been demonstrated (for example, *Opportunidades*, the urban version of Progresa, will not start with a randomised evaluation due to strong opposition to delaying the access of some people). This objection can be tackled at several levels. First, opposition to randomisation is less likely to succeed in an environment where strong support for it exists. This is the case particularly if the support is regulatory and prescribes that an evaluation is necessary before full-scale implementation. Second, if, as we have argued above, the evaluations are not financed by a loans but by grants, this may make it easier to convince partners of its usefulness, especially if it enables the country to expand a programme.

An example of such explicit partnership is a study on the effectiveness of HIV/AIDS education, currently being conducted in Kenya (Duflo, Dupas, Kremer and Sinei, 2003). With support from UNICEF, the government of Kenya has put together a teacher-training programme for HIV/AIDS education. For lack of funds, the coverage of the programme had remained very partial. The Partnership for Child Development, with grants from the World Bank, is funding a randomised evaluation of the teacher-training programme. ICS, a Dutch NGO, is organising training sessions with facilitators from the Kenyan government. The evaluation has made it possible to expand training to 540 teachers in 160 schools, which would not have been possible otherwise. The randomisation was not a ground for rejection of the programme by the Kenyan authorities. On the contrary, at a launch conference, Kenyan officials explicitly recognised the opportunity evaluation gave them of being at the forefront of efforts to advance knowledge in the area. The example of Progresa showed that government officials recognised the value of randomised evaluation and were actually prepared to pay for it. The very

favourable response to Progresa and the subsequent endorsement of the findings by the World Bank will certainly have an impact on how other governments think about experiments. Several examples of this kind could do much to change the culture.

4.3. NGOs are well-suited to conduct randomised evaluations, but will require technical assistance (for example, from academics) and outside financing

Governments are not the only vehicles through which randomised evaluations can be organised. Indeed, the evidence presented in this paper suggests that NGO projects can also be evaluated. Unlike governments, NGOs are not expected to serve entire populations. Even small NGOs can substantially affect budgets in developing countries. Given that many NGOs exist and that they frequently seek out new projects, it is often relatively straightforward to find NGOs willing to conduct randomised evaluations: hitches are more often logistical than philosophical.

For example, a set of recent studies conducted in Kenya was carried out through a collaboration with the Kenyan NGO, Internationaal Christelijk Steunfonds (ICS) Africa: ICS was keenly interested in using randomised evaluations to assess its impact, as well as to enable it to share credible evaluation results with other stakeholders and policymakers. A second example is the collaboration between the Indian NGO, Pratham, and MIT researchers, which led to the evaluation of the remedial education and computer-assisted learning programmes (Banerjee *et al*., 2003). This collaboration was initiated by Pratham seeking partners to evaluate their programmes; Pratham understood the value of randomisation and was able to convey the importance of such evaluations to the schoolteachers involved in the project.

However, while NGOs are well-placed to conduct randomised evaluations, it is less reasonable to expect them to finance them. The evaluations of the ICS de-worming programmes were made possible by financial support from the World Bank, the Partnership for Child Development, US National Institutes of Health (NIH), and the MacArthur Foundation. In the case of the Indian educational programmes, Pratham was able to find a corporate sponsor. India's second-largest bank, ICICI Bank, was keenly interested in evaluating the impact of the programme and helped to finance part of the evaluation. In general, given that accurate estimates of programme effects are international public goods, randomised evaluations should be financed internationally.

4.4. Costs can be reduced and comparability enhanced by conducting a series of evaluations in the same area

Once staff members are trained, they can work on multiple projects. Since data collection is the most costly element of these evaluations, crosscutting the sample can also dramatically reduce costs. For example, many programmes seeking to increase school participation were implemented in the same area and by the same organisation. Teacher incentive (Glewwe *et al*., 2003) and textbook (Kremer *et al*., 2002) programmes were evaluated in the same 100 schools in Western Kenya: one group had textbooks only, one had textbooks and incentives, one had incentives only, and one had neither. The effect of the incentive programme should thus be interpreted as the effect of an incentive programme conditional on half the schools having extra textbooks. Likewise in Vadodara, Pratham implemented a computer-assisted learning programme in the same schools where the remedial education programme evaluated by Banerjee and others (2003) was implemented. The programme was implemented only in Grade 4. Half the schools that had the remedial education programme in Grade 4 got the computer-assisted learning programme, and half the schools that did not have the remedial education programme got the computer-assisted learning programme. The preliminary results suggest that the effect on mathematics is comparable to the effect of the remedial education programme, but the cost is much smaller. Even holding constant the budget of process evaluation, a reallocation of part of the money currently spent on unconvincing evaluation would probably go a long way towards financing the same number of randomised evaluations. Even if randomised evaluations are revealed to be more expensive, the cost is likely to be trivial in comparison to the amount of money saved by avoiding the expansion of ineffective programmes. This suggests that randomised evaluation should be financed by international organisations, a point to which we will return below.

This technique must take into account potential interactions between programmes (which can be estimated if the sample is sufficiently large), and may not be appropriate if one programme renders the schools atypical. On the other hand, it enables us to evaluate the cost-effectiveness of different approaches to fighting a given problem. For example, several evaluations conducted in India in the same set of schools helped conclude that de-worming was the most cost-effective way of increasing school participation (compared to school uniform, school meals, or teacher incentives, Kremer,

2003), while merit-based scholarships are the most cost-effective way of increasing test scores (Kremer *et al*. (2004)).

4.5. The timing of evaluation and implementation

Prospective evaluations take time: convincing studies often last two or three years and it takes even longer to assess the long-term impact of the programme, which can differ substantially from the short-run impact . For example, Glewwe, Illias and Kremer (2003) suggest that a teacher incentive programme caused a short-run increase in test scores, but had no long-run impact, which they attributed to the practice of "teaching to the test". When the programme targets children but seeks to affect adult outcomes (which is the case for most education or health interventions), the delay between the programme and outcome may be very long. In these cases, it is not possible to wait for the impact before deciding whether or not to implement the programme.

While this is a real concern, it should not prevent the setting up of the evaluation on the first cohort exposed to the programme: while it is true that policy decisions will have to be taken in the meantime, it is surely better to know the impact at some point in the future. Moreover, it is often possible to obtain short-term results, which may give an indication of whether or not the programme could be effective, and may guide policy in the short run. For example, in the case of the evaluation of the HIV/AIDS teacher-training programme, an assessment was performed a few weeks after the programme was started (and while it was still ongoing). Students in the schools where the teachers were first trained were interviewed about whether HIV/AIDS was present in the curriculum in their school, and were administered a knowledge, attitude, and practice test. The preliminary results suggest that the programme was indeed effective in increasing the likelihood that HIV/AIDS is mentioned in class, and in improving students' knowledge about HIV/AIDS and HIV prevention. These results could be communicated immediately to the policymakers. The first result of an evaluation can also be combined with other results or with theory to provide an estimate of the final impact of the programme. Obviously, one has to be very careful about these exercises, outlining the purpose the evaluation results will serve, as well as the implications of assumptions. One should set up programmes to track long-run outcomes, which can then vindicate or invalidate these predictions. For example, Miguel and Kremer (2003) combined their estimate

of the impact of the de-worming programme on school participation on estimates of returns to education in Kenya to provide an estimate of the long-term impact on adult productivity. They then used this to construct cost-benefit estimates. They are also continuing to track the children exposed to de-worming drugs to directly estimate their long-run effect.

Finally, delaying expenditures may actually be worthwhile, particularly if this gives us an opportunity to learn more, given that we know so little about what does and does not work. After spending so many years funding education projects, it is disconcerting how little we know about programme effectiveness in education. In light of this, the fact that an evaluation takes two or three years (or even longer) seems a very short period of time. It may delay certain expenditure, but it will accelerate the process of learning how to apply funds usefully. The FDA requires randomised evaluation of drugs before they can be distributed. Occasionally, the delay it imposes on the approval of new drugs has created resentment (most recently, among associations representing AIDS victims). However, there is little doubt that randomised trials have played a key role in shaping modern medicine, and that they have accelerated the development of effective drugs.

4.6. Randomised evaluations have a number of limitations, but many of these limitations also apply to other techniques

Many of the limitations of randomised evaluations also apply to other techniques. In this subsection, we review four issues that affect both randomised and non-randomised evaluations (sample selection bias, attrition bias, spillover effects, and behavioural responses), and argue that randomised methods often allow for easier correction of these than non-randomised methods.

Sample selection problems could arise if factors other than random assignment influence programme allocation. For example, parents may move their children out of a school without a programme into a school with the programme. Conversely, individuals allocated to a treatment group may not receive the treatment (for example, because they decide not to take up the programme). Even if randomised methods have been used and the intended allocation of the programme was random, the actual allocation may not be. This problem can be addressed through “intention to treat (ITT)” methods

or by using random assignment. Although the initial assignment in this case does not guarantee that someone is actually either in the programme or in the comparison group, in most cases it is at least more likely that someone is in the programme group to which he or she was initially allocated. The researcher can thus compare outcomes in the initially assigned group and scale up the difference by dividing it by the difference in the probability of receiving the treatment in those two groups. In this way, the local average treatment effect estimate is obtained (Imbens and Angrist, 1994). Methods such as ITT estimates allow selection problems to be addressed relatively easily in the context of randomised evaluations, but it is often much more difficult to make these corrections in the case of a retrospective analysis.

A second issue affecting both randomised and non-randomised evaluations is differential attrition in the treatment and the comparison groups: those who participate in the programme may be less likely to move or otherwise drop out of the sample than those who do not. For example, the two-teacher programme analysed by Banerjee and others (2001) increased school attendance and reduced dropout rates. This means that when a test was administered in the schools, more children were present in the programme schools than in the comparison schools. If children prevented from dropping out by the programme are the weakest in the class, the comparison between the test scores of children in treatment and control schools may be biased downwards. Statistical techniques can be used to bound the potential bias, but the ideal is to try to limit attrition as much as possible. For example, in the evaluation of the remedial education programme in India (Banerjee *et al*., 2003), an attempt was made to track down *all* children and administer the test to them, even if they had dropped out of school. Only children who had left for their home village were not tested. As a result, the attrition rate remained relatively high but did not differ between the treatment and comparison schools, increasing confidence in the estimates.

Third, programmes may create spillover effects on people who have themselves not been treated. These spillovers may be physical, as found for the Kenyan de-worming programme by Miguel and Kremer (2003, forthcoming) when de-worming interferes with disease transmission and thus reduces worm infection both in children in the programme schools who did not receive the medicine and in children in neighbouring schools. Such spillovers might also operate through prices, as when the provision of school meals leads competing local schools to reduce school fees (Vermeersch, 2002).

Finally, there might also be learning and imitation effects (Duflo and Saez, forthcoming; Miguel and Kremer, 2003b).

If such spillovers are global (for example, due to changes in world prices), total programme impact will be difficult to identify with any methodology. However, if such spillovers are local, then randomisation at the level of groups can enable estimation of the total programme effect within groups and can generate sufficient variation in local treatment density to measure cross-group spillovers. For example, the solution in the case of the de-worming study was to choose the *school* (rather than the pupils within a school) as the unit of randomisation (Miguel and Kremer, 2003, forthcoming), and to look at the number of treatment and comparison schools within neighbourhoods. Of course, this requires a larger sample size.

One less-tractable issue is that the provision of inputs might temporarily increase morale among students and teachers, and hence improve performance. While this would bias randomised evaluations, it would also bias fixed-effect or difference-in-difference estimates. However, it is unclear how significant this issue is in practice, whereas, in contrast, we know that selection is a serious concern.

In summary, while randomised evaluation is not a bullet-proof strategy, the potential for biases is well known and can often be corrected. This stands in contrast to biases of most other types of studies, where the bias due to the non-random treatment assignments often cannot be assigned or estimated.

4.7. Publication bias appears to be substantial with retrospective studies; randomised evaluations can help address publication bias problems, but institutions are also needed

Publication bias is a particularly important issue that must be addressed. Positive results naturally tend to receive a large amount of publicity: agencies that implement programmes seek publicity for their successful projects, and academics are much more interested in, and able to publish, positive results than modest or insignificant results. However, many programmes clearly fail, and publication bias will be substantial if positive results are much more likely to be published. Available evidence suggests the publication bias problem is severe (DeLong and Lang, 1992) and especially significant with studies that employ non-experimental methods.

Publication bias is likely to be acute with retrospective studies. *Ex post,* the researchers or evaluators define their own comparison group, and thus may be able to select a variety of plausible comparison groups; in particular, researchers obtaining negative results with retrospective techniques are likely to try different approaches, or not publish. In the case of "natural experiments" and instrumental variable estimates, publication bias may actually more than compensate for the reduction in bias caused by the use of an instrumental variable. This is because these estimates tend to have larger standard errors, and because researchers looking for significant results will only select large estimates. For example, Ashenfelter and others (1999) show that there is strong evidence of publication bias in instrumental variables-based estimates of the returns to education: on average, the estimates with larger standard errors also tend to be larger. This accounts for most of the oft-cited result that instrumental estimates of the returns to education are higher than ordinary-least-squares estimates.

In contrast, randomised evaluations commit in advance to a particular comparison group: once the work to conduct a prospective randomised evaluation has been completed the results are usually documented and published even if the results suggest quite modest effects or even no effects at all.

As we will discuss below, it is important to establish institutions to ensure that negative results are disseminated. Such a system is already in place for medical trial results, and creating a similar system for documenting evaluations of social programmes would help to alleviate the problem of publication bias. Beyond allowing for a clearer picture of the effectiveness of interventions, this type of institution would provide the level of transparency necessary for systematic literature reviews to be less biased in their conclusions about the efficacy of particular policies and programmes.

4.8. Although any given randomised evaluation is conducted within a specific framework with unique circumstances, randomised evaluations can shed light on general issues

Without a theory of why a programme is effective, generalising from one well-executed randomised evaluation may be unwarranted. Similar issues of generalisability arise irrespective of the evaluation technique used. One way to learn about generalisability is to encourage adapted replications of randomised evaluations in

several different settings. It is possible that a programme that failed in one context might have succeeded in another, but adapted replications, guided by a theory of why the programme was effective, will go a long way towards alleviating this concern. This is one area where international organisations, which are already present in most countries, can play a key role. Such an opportunity was seized in implementing adapted replications of Progresa in other Latin American countries. Encouraged by the success of Progresa in Mexico, the World Bank encouraged (and financed) the implementation of similar programmes in neighbouring countries. Some of these programmes included randomised evaluations (for example, the *Programmea de Asignación Familiar* (PRAF) programme in Honduras), and are currently being evaluated with some variants that will help us better understand the impact of different rules of the programme.

Often the results of the first phase of a project may be difficult to interpret because of circumstances that are unique to this phase: a project may have failed as the result of implementation problems that could be avoided in later phases of the project; or, a project may have succeeded because it received more resources than a project in a more realistic situation or less favourable context. Even if the choice of the comparison and treatment groups ensures the internal validity of estimates, any method of evaluation is subject to external validity problems due to the specific circumstances of implementation. In other words, the results may not generalise to other contexts.

One problem specific to randomised evaluations is that members of either the treatment or comparison group could potentially change their behaviour, not due to the intervention, but due simply to the fact that they would know that they are part of a randomised evaluation. For example, provision of inputs might temporarily increase morale among beneficiaries and this could improve performance. Of course, to the extent that both groups change their behaviour in the same way, this will not lead to bias. It is also perhaps less likely that this will occur over a long period and that it will occur immediately after the introduction of the intervention. Some experimental designs can minimise the risk of such effects. For example, in Pratham's remedial education programme, *all* of the schools, but not all the grades, received the programme. It is, however, important to try to assess whether these effects are present. In his reanalysis of the project STAR data, Krueger (1999) exploits variation in class size within the control group occasioned by children's departure during the year to obtain a second estimate of the class size effect. By definition, this is not contaminated by John Henry

or Hawthorne effects, since all of the teachers in this sample belong to the control group. He finds no difference in the estimates obtained by these two methods.

Treatment effects may also be influenced by the scale of the programme. For example, the Columbian voucher programme analysed in Angrist and others (2002), described above, was implemented in a small-scale pilot. The rest of the school system remained unchanged however In particular, the number of students affected was too small to have an impact on the composition of the public and the private schools. If this programme were to be implemented on a large scale, it could influence the functioning of the school system, altering its impact (Hsieh and Urquiola, 2002). More generally, "partial equilibrium" treatment effects may be different from "general equilibrium" treatment effects (Heckman, Lochner and Taber, 1998). To address these problems, randomised evaluations performed at the "economy" level are necessary. This may be possible for programmes such as the voucher programme above, where the general equilibrium effects will plausibly take place at the level of the community, and where communities may or may not be randomly affected. It must be said, however, that we are not aware of an evaluation of this type.

One method of addressing questions about the external validity of any particular study is to implement adapted replications of successful (and potentially unsuccessful) programmes in different contexts. Such adapted replications have two advantages: first, in the process of "transplanting" a programme, circumstances will change and robust programmes will show their effectiveness by surviving these changes; second, obtaining several estimates in different contexts will provide some guidance about whether the programme has a notably different impact on different groups.

Two studies on school-based health interventions provide a good illustration. The first study (Miguel and Kremer, 2003) evaluated a programme of twice-yearly school-based mass treatment with inexpensive de-worming drugs in Kenya, where the prevalence of intestinal worms among children is very high. Seventy-five schools were introduced to the programme in random order. Health and school participation improved not only at programme schools, but also at nearby schools, due to reduced disease transmission. Absenteeism in treatment schools was 25% (or 7 percentage points) lower than in comparison schools. Including this spillover effect, the programme increased schooling by 0.15 years per person treated. Combined with estimates about the rates

of returns to schooling, the estimates suggest extremely high rates of returns of the de-worming intervention: the authors estimate that de-worming increases the net present value of wages by over $30 per treated child at a cost of only $0.49. One of the authors subsequently decided to examine whether these results generalised among pre-schoolers in urban India (Bobonis, Miguel and Sharma, 2002). The baseline revealed that, although worm infection is present, the levels of infection were substantially lower than in Kenya (in India, "only" 27% of children suffer from some form of worm infection). However, 70% of children had moderate to severe anaemia. The programme was thus modified to include iron supplements and was administered through a network of pre-schools in urban India. After one year of treatment, they found a nearly 50% reduction in moderate to severe anaemia, large weight gains, and a 7% reduction in absenteeism among 4- to 6-year-olds (but not for younger children). The results of the previous evaluation were thus by and large vindicated.[5]

Replication of the initial phase of a study in a new context does not imply delaying full-scale implementation of the programme if this is supported on the basis of existing knowledge. More often than not, however, the introduction of the programme can only proceed in stages, and the evaluation simply requires that participants be phased into the programme in random order. In addition, such adapted replications can be used to verify whether programme effects within samples change with covariance. For example, suppose that the effect of a given programme is smaller in schools with good teachers; one might consider whether in a different setting with much better teachers the effect would be smaller.

Institutions are required to provide incentives to conduct such replications and to aggregate the results to obtain a consistent picture of the impact of a particular approach.

It is worth noting that the exogenous variation created by randomisation can be used to help identify a structural model. Attanasio *et al*. (2001) and Berhman *et al*. (2002) provide two examples of using this exercise in combination with the Progresa data to predict possible effects of varying the schedule of transfers. For example, Attanasio

5. To make this point precisely, one would need a full cost-benefit analysis of both programmes to see whether the same improvement in human capital was achieved with the same expenditure. At this point, the paper on India does not yet have a cost-benefit analysis.

(2001) found that the randomised component of the Progresa data induced extremely useful exogenous variation that helped in the identification of a richer and more flexible structural model. These studies rest on assumptions that can be accepted or not, but at least they are freed from *some* assumptions by the presence of this exogenous variation.

The more general point is that randomised evaluations do not preclude the use of theory or assumptions: in fact, they generate data and variation that can be useful in identifying some aspects of these theories.

In order to give us guidance about key elements of a programme's success, we need a theory to predict the likely effectiveness of a specific programme. Importantly, theory will help unpack distinct components of a programme, and discriminate between significant and insignificant variants (Banerjee, 2002). For example, an economic analysis of the Progresa programme suggests that it may have been useful either because of its impact on income, or on women's bargaining power, or its effect on incentives. Aspects of the programme most likely to be relevant to the programme's success are the size of the transfer, its recipient, and the conditionality attached to it. In contrast, the colour of the food supplement distributed to the families, for example, is unlikely to be important. Replication of the programmes may then alter these aspects, to determine which is the most significant. This also suggests that programmes justified by some well-founded theoretical reasoning should be evaluated in priority, because the conclusions from the evaluation are then more likely to generalise.

Theory provides some guidance about what programmes are likely to work and, in turn, the evaluation of these programmes forms a test of the theory's prediction. Since prospective evaluations need to be planned in advance, it is also often possible to design pilot programmes in such a way that they assist in answering a specific question, or testing a specific theory. For example, Duflo (2003) reports on a series of randomised evaluations conducted in Kenya with Michael Kremer and Jonathan Robinson. They were motivated by the general issue of why so few farmers in one particular region of Kenya used fertilizer (only about 10%), despite the fact that it seemed to be profitable and was widely used in other developing countries, as well as in other regions of Kenya. They first conducted a series of trials on the farms owned by randomly selected farmers, and confirmed that, in small quantities, fertiliser is extremely profitable (the rates

of returns were often in excess of 100%). They then conducted a series of programmes to answer the following questions: do farmers learn when they try fertiliser out for themselves? Do they need information about returns or about how to use them? Does the experiment need to take place on their farm, or can it take place on a neighbour's farm? Do they learn from their friends? To answer these queries, the authors implemented several programmes. First, they randomly selected farmers to participate in the field trials, and followed their subsequent adoption, as well as that of a comparison group. Second, they also followed adoption of the friends and neighbours of the comparison farmers. Finally, they invited randomly selected friends of farmers participating in the trials to the important stages in the development of the experiment, and also monitored subsequent adoption. These questions are very important to our understanding of technology adoption and diffusion, and the ability to generate exogenous variation through randomised programme evaluation greatly helped in this understanding. Moreover, their answer also helped the NGO develop a school-based agricultural extension programme that aims to be both effective and cost-effective. A pilot version of this programme is currently being evaluated.

5. Roles for international agencies

5.1. Current practice

The examples discussed above illustrate that it is possible to obtain convincing evidence about programme impact by organising pilot projects to take advantage of the expansion of existing projects, or to take advantage of project design. While not all programmes can be evaluated using these methods, only a very small fraction of those that could be evaluated actually are. Most international organisations require that a fraction of programme budgets be spent on evaluation. Some countries also make evaluation compulsory (for example, evaluation of all social programmes is required by the Constitution in Mexico). However, in practice, this share of the budget is not always spent efficiently: evaluations are subcontracted to untrained consultancy outfits, with little guidance as to what they should achieve. Worse, they are sometimes entrusted to organisations that have an interest in the outcome, so the evaluators have a stake in the results they are trying to establish. When an evaluation is actually conducted, it is generally limited to a *process* evaluation: accounts are audited, the flows of resources

are tracked, the actual delivery of inputs is confirmed (for example, did the textbooks reach the school?) and qualitative surveys are used to determine whether the inputs were actually used by their beneficiaries (did the teachers use the textbooks?), and whether there is *prima facie* evidence that the programme beneficiaries were satisfied by the programme (were the children happy?). Process evaluation is clearly essential and should also be part of any programme evaluation: if no textbooks were actually distributed, that the programme has no impact is not surprising. However, only observing the beneficiaries' reactions to a programme can lead to very misleading conclusions about its effectiveness. Some programmes may, by all observations, appear to be resounding successes, even if they did not achieve their objectives. The emphasis on process evaluation implies that, more often than not, impact evaluations, when they take place, are an afterthought and are not planned from the programme's inception.

The District Primary Education Program (DPEP), the largest World Bank sponsored education programme, implemented in India, potentially offered very interesting evaluations, but these were jeopardised through lack of planning. DPEP was intended to showcase the ability to "go to scale" with education reform (Pandey, 2000). Case (2001) gives an illuminating discussion of the programme highlighting its features that rendered evaluation impossible. Seeking to improve the performance of public education, the DPEP was a comprehensive programme made up of teacher-training, inputs and classrooms. Districts were generally given a high level of discretion in how they chose to spend the additional resources.

Despite an apparent commitment to evaluating the programme carefully, several features made a convincing impact evaluation impossible. First, the districts were selected according to two criteria: low *level* of achievement (measured by low female literacy rates), but high *potential for improvement.* In particular, the first districts chosen to receive the programme were selected "on the basis of their ability to show success in a reasonable timeframe" (Pandey, 2000, quoted in Case, 2001). The combination of these two elements in the selection process suggests that any comparison between achievement levels in DPEP and non-DPEP districts would probably be biased downwards. Conversely, any comparison of improvements in DPEP and non-DPEP districts ("differences-in-differences") would probably be biased upwards. These fundamental constraints did not deter the DPEP from placing enormous emphasis on monitoring and evaluation: large amounts of data were collected, and numerous reports

were commissioned. However, the data collection process was conducted *only in DPEP districts*! This data can only be used for before/after comparisons, which clearly have little relevance in an economy undergoing rapid growth and transformation. If a researcher could even determine a credible identification strategy, he or she would have to use census or National Sample Survey (NSS) data.

5.2. The political economy of programme evaluation

We have argued that the omitted variable bias which randomised evaluations are designed to address is real and that randomised evaluations are feasible. They are no more costly than other types of surveys, and are far less expensive than pursuing ineffective policies. Why then, are they so rare? Cook (2001) attributes their rarity in education to the post-modern culture in US schools, which is hostile to the traditional conception of causation that underlies statistical implementation. Pritchett (2003) argues that programme advocates systematically mislead swing voters into believing exaggerated estimates of programme impact. Advocates block randomised evaluations since they would reveal the true impact of programmes to voters. Kremer (2003) proposed a complementary explanation, arguing that policymakers are not systematically fooled, but that they have difficulty gauging the quality of evidence, knowing that advocates can suppress unfavourable results. Programme advocates only present the highest estimates to policymakers, while opponents select the most negative estimates. Knowing this, policymakers rationally discount these estimates. For example, if advocates present a study showing a 100% rate of return, the policymaker might assume the true return is 10%. In this environment there is little incentive to conduct randomised evaluation. As the resulting estimates include no bias term, they are unlikely to be either sufficiently high or low for advocates to present them to policymakers. Even if results are presented to policymakers, those policymakers unable to gauge the quality of particular studies will discount them. Why would they fund a project that a randomised evaluation suggests has a 25% rate of return when advocates of competing projects claim a 100% rate of return?

In this context, international organisations can play a key role by encouraging randomised evaluations and funding them. Moreover, if it becomes easier for policymakers and donors to identify a credible evaluation when examples already exist (which seem plausible), this can actually launch a virtuous circle. Encouraging other

donors to recognise and trust credible evaluations will consequently encourage them to advocate these evaluations as opposed to others. In this way, they can contribute to creating a "climate" favourable to credible evaluation. The process of quality evaluation itself would then be scaled up above and beyond what the organisations can themselves promote and finance.

5.3. Evaluations in international organisations

International organisations could play several roles in promoting and financing rigorous evaluations.

It is almost certainly counterproductive to demand that *all projects* be subject to impact evaluation. Clearly, all projects must be monitored to ensure that they are actually implemented and to avoid misuse of funds. That said, some programmes simply cannot be evaluated with the methods discussed in this paper and, even among projects that could potentially be evaluated, not all would benefit from it. In fact, the value of a poorly identified impact evaluation is very low and its cost, in terms of credibility, is high. This is particularly the case if international organisations take a leading role in promoting quality evaluation. A first objective is thus to cut down on the number of wasteful evaluations; any proposed impact evaluation should be reviewed by a committee before money is spent on data collection. The committee's responsibility would be to assess the ability of the evaluation to deliver reliable causal estimates of the project's impact. A second objective would be to conduct credible evaluations in key areas. In consultation with a body of researchers and practitioners, each organisation should determine key areas in which it will promote impact evaluation. Randomised evaluations could also be established in other areas when the opportunity occurs.

Credible impact evaluations require a great deal of work and, in addition, the benefits of credible impact evaluations extend far beyond the organisation conducting the evaluation. These factors mean that incentives to conduct rigorous evaluations are less than socially optimal. One promising remedy is to embed structures within the institutional framework of international agencies that will provide sufficient incentives for evaluators. Given the current scarcity of randomised evaluations within the institutional environment of international organisations, there may be scope for setting up a specialised fund to encourage, conduct and finance rigorous impact evaluations, and

to disseminate the results. As we will briefly discuss below, the potential for such a fund is tremendous: there exists a ready-made potential supply of evaluators both within the international agencies themselves, as well as within academia. In addition, collaborations with NGOs offer many opportunities for evaluating policies of broad relevance.

An evaluation fund such as this would encourage data collection and the study of true "natural randomised evaluations" with programme-induced randomisation. As we mentioned in Section 2 above, randomised evaluations are not the sole method of conducting good impact evaluations. However, other types of evaluations are conducted much more routinely in comparison to randomised evaluations, given both their value and the many opportunities for carrying them out. Part of the problem is that responsibility for conducting these evaluations belongs to no one organisation or individual, and hence there is insufficient investment in conducting them. In addition, because there are common features to all evaluations, a specialised unit with specific expertise could assist enormously. Since impact evaluations generate international public goods, the unit should have a budget that would be used to finance and conduct rigorous evaluations of internal and external projects. This unit should also conduct its own evaluation projects in the key areas identified by the organisation.

As previously discussed, the unit should also work with partners, especially NGOs and academics. For external projects, a committee within the unit (potentially assisted by external reviewers) could receive proposals from within the organisation or from outsiders, and use this as a basis for choosing which projects to support. The unit could also encourage replication of significant evaluations by sending out calls for specific proposals. The project could then be conducted in partnership with others from the unit or researchers (academics, in particular). The unit could provide both financial and technical support for the project, with dedicated staff and researchers. Over time, on the basis of the acquired experience, the unit could also serve as a more general resource centre by developing and diffusing training modules, tools and guidelines (survey and testing instruments, as well as software that can be used for data entry and to facilitate randomisation, similar in spirit to tools produced by other units in the World Bank). The unit could also sponsor training sessions for practitioners.

Another role the unit could serve, once it had established a reputation for quality, is that of a dissemination agency (a "clearing house" of some sort). To be useful,

evaluation results must be accessible to practitioners both within and outside development agencies. A key role of the unit could be to conduct systematic searches for all impact evaluations, assess their reliability, and publish the results in the form of policy briefs and in a readily accessible searchable database. The database would ideally include all information that could be useful in interpreting the results (estimates, sample size, region and time, type of project, cost, cost-benefit analysis, caveats, and so forth), as well as references to related studies. The database could include both randomised and non-randomised impact evaluations satisfying given criteria, provided that the different types of evaluation are clearly labelled. Evaluations would need to satisfy minimum reporting requirements to be included in the database, and all projects supported by the unit would have to be included in the database, whatever their results.

As previously discussed, such a database would help alleviate publication bias, which may be substantial if positive results are more likely to be published. Academic journals may not be interested in publishing the results of failed programmes, but from the policymakers' point of view, knowledge about negative results is just as useful as knowledge about successful projects. Comparable requirements are placed on all federally funded medical projects in the US. Ideally, over time, the database would become a basic reference for organisations and governments, especially as they seek funding for their projects. This database could kick-start a virtuous circle, with donors demanding credible evaluations before funding or continuing projects, more evaluations being conducted, and the general quality of evaluation work rising.

6. Conclusion: using evaluation to build long-term consensus for development

Rigorous and systemic evaluations could leverage the impact of international organisations well beyond their simple ability to finance programmes. Credible impact evaluations are international public goods: the benefits of knowing that a programme works or does not work extend well beyond the organisation or the country implementing the programme. Programmes that have been shown to be successful can be adapted for use in other countries and scaled up within countries, while unsuccessful programmes can be abandoned. Through promoting, encouraging and financing rigorous evaluations

(such as credible randomised evaluations) of the programmes they support, as well as of programmes supported by others, international organisations can provide guidance to other donors, governments and NGOs in the ongoing search for successful programmes. In this way, they will improve the effectiveness of development aid. Moreover, by credibly establishing which programmes do and do not work, the international agencies can both counteract scepticism about spending aid effectively and build long-term support for development. In this, randomised evaluations provide an opportunity to achieve a real "scaling up".

REFERENCES

Angrist, J., E. Bettinger, E. Bloom, E. King and M. Kremer (2002), "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment" *American Economic Review* 92 (5): pp. 1535-58.

Angrist, J. and A. Krueger (1999), "Empirical Strategies in Labor Economics" in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics, Vol. 3A*. Amsterdam: North Holland: pp. 1277-1366.

Angrist, J. and A. Krueger (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments" *Journal of Economic Perspectives* 15 (4): pp. 69-85.

Angrist, J. and V. Lavy (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics* 114 (2): pp. 533-575.

Ashenfelter, O., C. Harmon, and H. Oosterbeek (1999), "A Review of Estimates of the Schooling/ Earnings Relationship, with Tests for Publication Bias" *Labour Economics* 6 (4): pp. 453-70.

Attanasio, O., C. Meghir and A. Santiago (2001), "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresa" mimeo, Inter-American Development Bank.

Banerjee, A. (2002), "The Uses of Economic Theory: Against a Purely Positive Interpretation of Theoretical Results" mimeo, MIT.

Banerjee, A., S. Cole, E. Duflo and L. Linden (2003), "Remedying Education: Evidence from Two Randomized Experiments" mimeo, MIT.

Banerjee, A. and R. He (2003), "The World Bank of the Future" *American Economic Review, Papers and Proceedings* 93 (2): pp. 39-44.

Banerjee, A., S. Jacob and M. Kremer (with J. Lanjouw and P. Lanjouw) (2001), "Promoting School Participation in Rural Rajasthan: Results from Some Prospective Trials" mimeo, Harvard-MIT.

Behrman, J., P. Sengupta and P. Todd (2002), "Progressing Through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Mexico" mimeo, University of Pennsylvania.

Bertrand, M., E. Duflo and S. Mullainathan (2003), "How Much Should We Trust Difference in Differences Estimates?", forthcoming in *Quarterly Journal of Economics*.

Besley, T. and A. Case (2000), "Unnatural Experiments? Estimating the Incidence of Endogenous Policies,' *Economic Journal* 110 (467): F672-F694.

Bobonis, G., E. Miguel and C. Sharma (2002), "Iron Supplementation and Early Childhood Development: A Randomized Evaluation in India" mimeo, University of California, Berkeley.

Buddlemeyer, H. and E. Skofias (2003), "An Evaluation on the Performance of Regression Discontinuity Design on Progresa", Institute for Study of Labor, Discussion Paper No. 827.

Campbell, D. T. (1969) "Reforms as Experiments", *American Psychologist,* 24: pp. 407-429.

Card, D. (1999), "The Causal Effect of Education on Earnings" in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics, Vol. 3A,* Amsterdam: North Holland, pp. 1801-63.

Case, A. (2001), "The Primacy of Education" mimeo, Princeton University.

Chattopadhyay, R. and E. Duflo (2001), "Women as Policy Makers: Evidence from an India-Wide Randomized Policy Experiment", NBER Working Paper 8615.

Chattopadhyay, R. and E. Duflo (2003) "Women as Policy Makers: Evidence from an India-Wide Randomized Policy Experiment", mimeo, MIT.

Cook, T. D. (2001) "Reappraising the Arguments Against Randomized Experiments in Education: An Analysis of the Culture of Evaluation in American Schools of Education", mimeo, Northwestern University.

Cronbach, L. (1982), *Designing Evaluations of Educational and Social Programmes,* San Francisco, Jossey-Bass.

Cronbach, L., S. Ambron, S. Dornbusch, R. Hess, R. Hornik, C. Phillips, D. Walker and S. Weiner (1980), *Toward Reform of Programme Evaluation,* San Francisco: Jossey-Bass.

Cullen, J. Berry, B. Jacob and S. Levitt (2002), "Does School Choice Attract Students to Urban Public Schools? Evidence from over 1,000 Randomized Lotteries", mimeo, University of Michigan.

DeLong, J. B. and K. Lang (1992), "Are All Economic Hypotheses False?", *Journal of Political Economy* ,100 (6): pp. 1257-72.

Duflo, E. (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", *American Economic Review,* 91 (4): pp. 795-813.

Duflo, E. (2003), "Poor but Rational?", mimeo, MIT.

Duflo, E., P. Dupas, M. Kremer and S. Sinei (2003), "Evaluating HIV/AIDS Prevention Education in Primary Schools: Preliminary results from a randomized controlled trial in Western Kenya", mimeo, Harvard-MIT.

Duflo, E. and M. Kremer (2003), "Use of Randomization in the Evaluation of Development Effectiveness", mimeo, MIT.

Galasso, E., M. Ravallion and A. Salvia (2002), "Assisting the Transition from Workfare to Work: A Randomized Experiment" mimeo, Development Research Group, World Bank.

Gertler, P. J. and S. Boyce (2001), "An Experiment in Incentive-based Welfare: The impact of Progresa on health in Mexico" mimeo, University of California, Berkeley.

Glewwe, P., Nauman I. and M. Kremer (2003), "Teacher Incentives", mimeo, Harvard University.

Glewwe, P., M. Kremer, S. Moulin and E. Zitzewitz (2003), "Retrospective vs. Prospective Analyses of School Inputs: The case of flip charts in Kenya", forthcoming in *Journal of Development Economics.*

Heckman, J., L. Lochner and C. Taber (1998), "General Equilibrium Treatment Effects: A Study of Tuition Policy", NBER Working Paper 6426.

Hsieh, C.-T. and M. Urquiola (2002), "When Schools Compete, How Do They Compete? An assessment of Chile's nationwide school voucher programme", mimeo, Princeton University.

Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica,* 62 (2): pp. 467-475.

Kremer, M. (2003), "Randomized Evaluations of Educational Programmes in Developing Countries: Some Lessons", *American Economic Review Papers and Proceedings*, 93 (2): pp. 102-115.

Krueger, A. (1999), "Experimental Estimates of Education Production Functions", *Quarterly Journal of Economics*,114 (2): pp. 497-532.

LaLonde, R. (1986), "Evaluating the Econometric Evaluations of Training with Experimental Data", *American Economic Review,* 76 (4): pp. 604-620.

Meyer, B. D. (1995), "Natural and Quasi-experiments in Economics", *Journal of Business and Economic Statistics,* 13 (2): pp. 151-161.

Miguel, E. and M. Kremer (2003), "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities", forthcoming in *Econometrica*.

Morduch, J. (1998), "Does Microfinance Really Help the Poor? New evidence from flagship programmes in Bangladesh", mimeo, Princeton University.

Narayanan, D. (ed.) (2000), *Empowerment and Poverty Reduction: A Sourcebook*, World Bank, Washington DC.

Pandey, R. S. (2000), *Going to Scale With Education Reform: India's District Primary Education Programme, 1995-99,* Education Reform and Management Publication Series, Volume I, No. 4., World Bank, Washington DC.

Pitt, M. and S. Khandker (1998), "The Impact of Group-Based Credit Programmes on Poor Households in Bangladesh: Does the Gender of Participants Matter?", *Journal of Political Economy,* 106 (5): pp. 958-996.

Pritchett, L. (2003), "It Pays to be Ignorant: A Simple Political Economy of Rigorous Programme Evaluation", forthcoming, *Journal of Policy Reform*.

Rosenbaum, P. R. (1995), "Observational Studies", in *Series in Statistics,* New York; Heidelberg; London: Springer.

Shultz, T. P. (2001), "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Programme", forthcoming, *Journal of Development Economics*.

Comment

by
Hans-Rimbert Hemmer
University of Giessen / Germany and Kreditanstalt für Wiederaufbau (KfW), Frankfurt

1. Introduction

In practical development co-operation, a strong tendency can be observed to shift aid more and more away from isolated projects to sector programmes. Accordingly, programme evaluation instead of project evaluation is necessary for finding out how successful these programmes are. The methodology of evaluating these sector programmes, however, has not yet reached the same qualitative level as project evaluation methodology did. Insofar it is important to direct more research activities to the field of sector evaluation methodology. Esther Duflo's paper is, in this context, a highly welcomed contribution to sector programme evaluation methodology.

The basic principle of evaluation is the "with-and-without" comparison. But what is the "without case"? In development practice, here we heavily depend on speculation. The question arises as to what extent the speculative character of the "without scenario" taken as benchmark can be reduced and the objectivity of the evaluation be increased. Duflo's paper is a remarkable (and, as I want to add at the very beginning of my comments, successful) attempt to contribute to a better identification of what can be taken as a proper "without scenario". She proves her high competence in evaluation methodology, as well as her profound knowledge of the state of the evaluation art. Her presentation of the concept of prospective randomised evaluations and its comparison with the "competing" methods of propensity score matching, "difference-in-difference" techniques, and regression discontinuity design is convincing, logically sound and a real pleasure to read.

According to my own knowledge and understanding, the two most important problems inherent in Esther's preferred methodology of prospective randomised evaluations are:

a) the selection of a credible comparison group whose development can be taken as the "without-scenario development", and
b) the identification of the objective function that is behind the impact measurement.

2. Selection of a credible comparison group

Esther shows what steps have to be taken in order to find a valid result. All the points made by her are correct and necessary at the same time. Nevertheless, one aspect cannot be isolated with her methodology: the larger a programme, the greater the probability that the comparison group already knows about it. In such a case, it cannot be excluded that the mere knowing of the existence of such a programme (and, maybe, also of some cornerstones of it) already influences the behaviour of the comparison group in a way that the development process of the comparison group also reflects to a certain extent the programme itself. If the programme is valued positively by the members of the comparison group, they might be tempted to influence the evaluation result positively—that means in a way that evaluation shows the good impact of the programme—with the implicit expectation that the programme will be extended to other groups (and, hopefully, also to the comparison group). On the contrary, if the programme is valued negatively, there might be the temptation to "proof" an only marginal impact of the programme in order to have it stopped and/or to prevent its extension to the comparison group.

The larger a programme is designed and the better it is known, the greater is the danger of an "impact bias" in evaluation by comparing the programme consequences for the target group with a control group—irrespective how this control group has been selected. The smaller the programme, the lesser is this danger—but then the closer the programme may come to a single project.

Let me add another question related to these points: how should we deal with policy interventions caused by the programme (be it a precondition for the programme itself

or be it a reaction to some programme results) which have impact on other people outside the programme target group? Does this mean that we cannot find a real "without scenario" at all? Or does it limit strongly the applicability of the methodology proposed by Esther to programmes other than those presented to us as examples?

3. Choice of the objective function

The second problem I want to touch on is the identification of the objective function that is behind the impact measurement. Esther is in favour of regression techniques to find the impact—an approach I feel many sympathies for. But regression analysis is linked to regressible effects! According to my own experience as a project evaluator, I know the huge deficiencies evaluators are normally confronted with if they try to find the real objective function behind the programme. Many times, only some objectives are documented, and very often qualitative objectives play a crucial role, just to mention two aspects. In both cases, the impact of the programme cannot be measured at all. It begins with the inability to identify how much objective fulfilment has been reached, continues with the inability to discriminate between programme effects and autonomous effects (due to the shortages of the comparison with a control group), and finally, in the case of a complex objective function with a subdivision in general objectives and sector programme targets (that are supposed to positively contribute to the general objectives), there may be conflicting partial impacts that cannot be compensated.

In addition to that, it cannot be excluded that the objectives of the programme may be acceptable for the target group but not for the comparison group, once again with the consequence of objective-induced behavioural differences.

Comment

by
Elisabeth Sadoulet
University of California at Berkeley

Esther should be congratulated for providing us with an excellent review of the state of the arts in impact analysis of development programmes. It benefits from the considerable experience she has with the subject and from her own contributions to the field, both methodological and with the practice of impact analysis. Her comments on the advantages of a randomisation approach, as well as on the difficulties with implementation of this approach and on recommendations as to how to overcome them, are particularly relevant to this meeting.

I will confine my comments to six observations that complement Esther's presentation. They derive from the experience I have had with impact analysis of such programmes as Progresa and the Micro-regions Strategy in Mexico, Bolsa Escola in Brazil, microfinance projects in Guatemala, and support to producers' organisations in Burkina Faso and Senegal. These comments stress the following:

(1) the need to transform impact analysis from an international public good to a private good with large spillover effects, thus justifying budgeting the cost of the evaluation, at least partially, into the project;
(2) the need to be clear about project objectives against which the evaluation is made, particularly when projects are participatory, and when objectives become an endogenous choice of participants;
(3) the need to broaden the dimensions of experimentation of projects, and to proceed in a sequential Bayesian fashion in answering an array of questions;

(4) the important role of calibrating structural models on pilot experiments to broaden the dimensions of simulation to instruments and contexts that could not possibly be done experimentally;
(5) the important and under-used potential of using administrative data for both prospective and retrospective evaluations of programmes;
(6) the need to endogenise the political economy of project implementation as a key dimension of success.

1. From international public goods to private goods with large spillover effects

As an international public good, impact evaluation will always be severely under-funded, in spite of exhortations by academics and global social planners that such investments can have handsome payoffs. The main reason is the inescapable logic of free-riding. This can be seen in the World Bank's own practice. Large projects such as support to the land reform in Brazil, which absorbed $2-3 billion in loans, or to CDD programmes, which currently absorb some $3 billion annually, remain un-assessed.

The other reason why impact evaluation of programmes is looked on as an international public good is that results from impact evaluations generally come too late for use by the project. Hence, it is not surprising that project managers are not interested in paying the cost of evaluation. Even country ministers will often say that they are not interested in funding project evaluation because their political time is shorter that the time span needed for evaluators to obtain results.

Two changes need to be made to internalise some of the benefits of evaluation in the project, and hence create incentives for at least partial funding of the evaluation by the project itself:

(1) *Just-in-time delivery of results*

Evaluation has to be designed so that short-term results are available and delivered to project managers while the project is still active. This requires caution (as discussed by Esther) that short-run outcomes not be confused with longer-term outcomes. In some

cases, achieving these early outcomes may be at the cost of worse performance in the longer run. However, as Esther mentions, there are carefully chosen short-term indicators of impact that can usually be defined. And this is easier for some projects, like remedial education, than for other projects that aim at raising incomes or reducing environmental degradation.

(2) *Evaluation as part of results-based management*

Impact evaluation should be conceived with a double purpose:

- impact evaluation for accountability to outside sponsors. This is the traditional objective of evaluation. To fulfil this objective, evaluation has to be done by outside auditors and all data used must be verifiable by third parties;
- impact evaluation as part of a participatory learning process leading to change. For this, evaluation should be part of a participatory process engaging members of the organisation who contribute information on indicators of success and failure, and internalise results from impact analysis in the design and practices of the organisation.

These two objectives are not incompatible, but they have rarely been implemented jointly. Doing so requires engaging both programme personnel and external auditors in the evaluation, and making sure that the accountability purpose is not being perverted by the management function. Yet, success in managing evaluation in this fashion allows benefits to be delivered to the organisation, and hence part of the costs of the evaluation to be internalised as a private good in the project's operational budget.

2. Definition of programme objectives as evaluation criteria

Participatory projects have many advantages. They make it possible to access local information and to mobilise local social capital. They often give themselves the objective of empowerment of participants through the participatory process itself. Well-known examples are the Community Driven Development and the Village Level Participatory Approaches. In this case, delegation to the agents of the choice of goals is part of the approach.

Which goals are to be used for evaluation: those of the external agency, or those set for themselves by community members? There is a genuine dilemma in letting the community members set the goals for themselves and then use those same goals for a performance evaluation, as different communities set goals for themselves unevenly difficult to reach. But, worst is when the external evaluator sets *ex-post* the evaluation criteria that do not correspond to those of the agents in the programme.

The difficulty arises also for programmes with multiple objectives. Even something as simple as conditional cash transfers are evaluated as transfers by some (impact on poverty reduction) and as price effects through the condition by others (impact on educational achievements).

3. Broadening the dimensions of the evaluation

Experimentations in randomised trials such as Progresa are too narrow: they only concern inclusion/exclusion, and not the terms of inclusion. Important dimensions about which there is need for experimentation in conditional cash transfer programmes include: level of the transfer for optimal uptake, choice of eligible households, community vs. individual targeting, gender of the transfer recipient, complementary programmes to maximise the multiplier effects of transfers, etc.

Since we typically lack degrees of freedom for experimenting in multiple dimensions, we need to proceed in a sequential Bayesian fashion, the way pharmaceutical companies proceed in experimenting with new drugs: raise a question, proceed with experimentation, and move on to the next question as soon as a statistically valid answer has been obtained to the question raised. It is quite likely, for instance, that there is no need to proceed with three years of experimental transfers of a set amount to measure impact on school enrolment. At the same time, moving on to other dimensions of experimentation as soon as we have enough evidence on the initial question raised makes it possible to calibrate the programme and make it eventually much more efficient. In the case of Progresa, for instance, it should have been almost immediately obvious that transfers are not justified to increase enrolment in primary school. In addition, experimenting with alternative targeting rules (e.g., focusing on children most at risk of not going to school, such as those living far from school) and with calibration

of transfers according to the situation of the child in order to maximise impact can have very large efficiency effects, as we have shown for the case of Progresa. Some of these effects can be derived *ex-post* from specific features in programme implementation (such as the existence of a household cap in Progresa transfers that results in unequal transfers across children), but, in general, well-designed experimentation is needed.

4. Calibration of structural models

Even if we broaden the range of questions raised for experimentation, and sequence their implementation for cost effectiveness, there are too many dimensions of interest to attempt to answer them all experimentally. Use of structural models that capture behavioural responses to transfers according to context specificity (including presence of complementary programmes) remains largely underused. These models should be calibrated on pilot experiments, specifically designed for that purpose. The models are the only available tool to capture the many dimensions of heterogeneity, as well as long-term and general equilibrium effects.

As an example, behavioural models we constructed to represent household responses to Progresa make it possible to explore the role of complementary programmes on educational attainments, such as remedial education and information on job opportunities for graduates. These complementary interventions are not in the current experimental design, and their impact on schooling decisions could consequently not be derived from reduced form models.

5. Using administrative data

Data collection remains a very costly and time-consuming undertaking, limiting the number, scope and usefulness of impact evaluations. However, in many situations, data can be collected through the programme itself, particularly when it is engaged in results-based management. We are working in this fashion with several commercial micro-lenders in Guatemala, using their administrative records as our database for impact analysis of reforms they introduced, such as use of credit bureau records for client screening. Participating institutions can be asked to modify the information they keep

on former or rejected clients and to complement the array of exogenous variables they collect to characterise clients. Mutually beneficial arrangements can be defined with participating institutions, particularly if results can be delivered on a just-in-time basis and support results-based management. These data files can be massive and give a time dimension that could never be achieved by survey data. Donors are well-placed to negotiate with beneficiary institutions the keeping of and access to administrative data. Massive administrative data are also available through local levels of governance that need to satisfy central government demands for transparency. Brazil is a good illustration of this, with most of these data available on-line to meet Federal government requirements.

6. Endogenising the political economy of project implementation

Many projects are implemented through decentralisation to local levels of governance, and through the active participation of agents. This is because much private information is public at the local level, local social capital can be mobilised for enforcement as a consequence of interlinked transactions, and project accountability with locally elected officials can expectedly be greater. It would be naïve, however, to evaluate impact without endogenising the political economy processes that condition outcomes. This requires a good understanding of local institutions and local political processes, of how decisions are made, how influence is exercised, and how clientelism and vote-buying enter into the appropriation of project rents. Data need to be collected for this purpose at the level of decision-making institutions (municipalities, local councils, interest groups). In Brazil, Bolsa Escola/Bolsa Familia delegates the selection of beneficiaries to municipalities and their local councils. Impact of the programme is thus mediated by local political economy processes, leading to heterogenous outcomes according to the local polity. With 5,300 municipalities engaged in this process, this constitutes an incredible laboratory for observing local political economy processes in action. Understanding their role is an integral component of the impact analysis of development projects.

Concluding remarks

by
Jean-Michel Severino
Director General of the Agence Française de Développement

Ladies and Gentlemen,

For a development agency, no subject is more important than aid effectiveness. This lies at the very heart of our existence and being incapable of responding to such an issue would be quite unthinkable. Paradoxically, however, the deeper we delve into this subject, the harder it becomes to hang on to our certainties. Somehow, research efforts leave the aid practitioner facing even greater contradictions and difficulties than before. Yet this process or questioning is, of course, a productive one.

While listening to the conference speakers, papers and reactions, as someone who is part of a development agency, one basically says to oneself: *"Conditionality does not work. Aid is effective for reasons that do not correspond to those we would wish for. Performance-based allocation has serious limits, but other methods are also problematic. Aid reduces countries' fiscal efforts and does not affect the global volume of public spending."* To exaggerate slightly, I would say that the only reassuring thing I have learned from reading the papers is that the CPIAs (*Country Policy and Institutional Assessments*), used by the World Bank to judge a country's "performance", are not worth very much. I would also agree with the critical charge brought by Ravi Kanbur. As a World Bank executive, I once had to manage these famous CPIAs, but now the abyss of perplexity they threw me into, the incoherence and sheer illegitimacy of their substance have finally been brought to light, with all the academic authority they deserve. I hope that, after

Ravi's intervention, the nature of the debate on performance-based allocation criteria will change and that we will finally escape the world of ideology.

I would, however, like to say that all this is no cause for despair—here in the 12th *arrondissement* of Paris, or on 19th Street in Washington, or wherever else people are supposed to be interested in the fate of humanity—and I would briefly like to explain why. As I have fortunately been given some time to speak, I shall take this opportunity to tell you why the debate on conditionality seems both a core issue and, yet, one poorly addressed by the scientific community, forcing operationals into a questioning that I feel is somewhat perverse. I shall then conclude with an appeal for help for the years to come.

To start with the first point. Why is it that all these doubts and questions on aid effectiveness do not sap the efforts of agencies like ours to tackle these subjects? And, here, I would like to share three thoughts.

First, basically all that has been said—for instance, by Jan Gunning—about "why give aid?", with respect to evaluating what we do, is utterly fair. For institutions like ours, the choice of good strategic positioning is absolutely fundamental. AFD's strategic orientation plan has defined our positioning as being where the interests of South and North countries meet, that is to say, very broadly where issues concerning the fight against poverty and global public goods converge. This is absolutely essential, because, in the long run, you cannot convince French tax-payers to intervene in development aid issues solely on the basis of compassion or global redistribution, without risking some kind of backlash. The question of global public goods allows the question of reciprocity and convergence of interests to be raised, without denying poverty issues or falling into the trap of a trade backlash—or its operational consequence: tied aid.

Obviously, this "double dividend" theory (global public goods and the fight against poverty) only really functions if there is additional aid flow without excessive fungibility. Otherwise, official development aid (ODA) interventions would dissolve into the local budgetary and macroeconomic reality. However, even were there no additionality and even if fungibility were extreme—which I do not believe—this positioning at least allows for good communication on aid policy, with a more solid grounding. This brings us back to the distinctions made by Jan Gunning this morning.

In this sense, the whole discussion on issues of fungibility and additionality, along with the consequences that can be drawn, leads aid practitioners to explore more thoroughly how they conceive interventions. Some time ago, in a meeting at the CERDI, we noted that there is actually less fungibility than macroeconomists think, and more than those working in aid believe. I certainly believe that instruments and methods allowing us to tackle different contexts do exist. But I do not wish to prolong the debate on this point right now.

A second remark: the international community says that it is intent on bettering its effectiveness, when, in fact, what it is really striving to improve is its efficiency. When we look at the contents of discussions on aid practices, we see that a great deal of effort is spent on problems of cost-effective aid management, rationality, good harmonisation, coherence, etc. These efforts, however, do not target effectiveness, but rather efficiency. Pointing this out does not imply detracting from their value, as these efforts are absolutely essential. Why so?

If we take the debate on aid motivations, a great deal is said about our profession's basic missions: the fight against poverty, geostrategic aspects, trade, etc.—there is a whole spectrum of terms. Yet, I think that we often underestimate the fact that, for those who mark out the posts in this policy area (our leaders, the politicians, the G7, G8 or OECD heads of state, the parliaments), aid's basic legitimacy and grounding is also embodied in the aid relationship itself. Let me try to explain.

If we examine the French President's statements (here, I should like to make a very modest analysis) on the legitimacy and rationality of official development aid, we see that his presentation focuses entirely on the international-level political and social risks that could arise, were sufficient generosity and solidarity not forthcoming. This is the fundamental political logic of everything we do, and the basis on which he arbitrates, year after year, to ensure our supplementary budgets. Underpinning this motivation and, of course, beyond the desire to change the world's economic and social balances, there is also the notion that aid's *raison d'être* is to show its beneficiaries the very fact that it exists. Thus, if what is fundamental is the relationship itself, many of the questions we ask ourselves about effectiveness become entirely secondary. This is not so, for the question of efficiency. Certainly, the taxpayer is not indifferent to knowing whether this same level of generosity, solidarity, political impact, etc. can be obtained for 1, 1/10,

1/100 or 1/1000. The international community must therefore work on the efficiency issue very thoroughly.

The search for how best to identify impacts is not totally devoid of interest either. Official development aid has a wide diversity of audiences. One of the problems here is the fact that ODA policy is fragmented and dispersed. Certainly, there are centralised donors who allocate budgets (donor countries' political leadership), but they are surrounded by a host of lobbies —professional, corporatist or individuals pursuing public-policy or interest-based objectives. For many of these groups, impact evaluation measurements —whether they concern issues of global public goods, poverty, trade or global strategy—are very important. They see ODA professionals as accountable and duty-bound to set up tools to measure their performance.

This remark, as you will immediately realise, highlights the fact that there can be no central evaluation function or single type of impact measurement. Conversely, there are a wide number of impact and evaluation measures depending on the group that, *in fine*, commissions (explicitly or implicitly) the policies we implement. Even if, *in fine*, some of these groups did not demand account of these impacts, there is certainly a strong logic—in terms of capacity improvement and support to the countries where we intervene—in favour of organising these evaluation and impact measurements, since they help to drive improvements in countries' statistics and evaluation systems. When we measure our own performance as an ODA agency or system, we are at the same time evaluating our partners' performance; they thus benefit from the efforts that we are making alongside them.

Now, let me share with you a few personal thoughts about the question of optimal allocation and conditionality. These are fundamental because, for some years now, we have been living under a sort of imperialism of "resource allocation according to the performance of beneficiary countries". This widespread intellectual concept, which is now predominant in many aid agencies and among the international intellectual community, would pose less of a problem, were it not based on two fundamental assumptions, presented as research results, and which you have debated today:

- first, donors are incapable of achieving policy changes in the countries where they intervene;

- second, in recipient countries, the global volume of public expenditure devoted to development issues cannot be modified.

These principles give equal support to the theory of aid allocation grounded on performance measured from past results (for example, as with Ravi Kanbur's proposals), and that using the CPIA-based performance criteria. Yet, these analyses possess three basic weaknesses that make them very difficult to apply and of no use in governing our interventions. The first weakness is that, in all the studies I have just alluded to, ODA is generally seen as a comprehensive undertaking that should impact economic growth. Yet, what really happens with ODA budgets (those taken into account in all the econometric regressions) has little to do with this intended objective. If we look at French development aid: a sizeable share of some 6.5 billion is linked to debt cancellation; another to crisis management; another (very substantial in France) to the cost of taking students from developing countries into French universities; another share goes to cultural policy and scientific and technical co-operation (which is very indirectly linked to economic growth issues); and yet another part is allocated to the very substantial research budget, (particularly, for the IRD and CIRAD). And so it goes on… When we come to the only free and programmable part of the aid, a considerable portion is earmarked for technical assistance, of which a good part is directed towards actions relating to political influence and sovereignty in foreign countries.

It is relatively difficult to reconcile aid's global image with these very diverse aims and points of action. Here, I would like to add that within ODA, we have no problem mixing the apples and oranges of subsidies and subsidised loans, the budgetary costs of which are attributed according to very specific conventions – which I will not come back to.

In short, in the notion of "ODA", we are using a category which, quite frankly, has no homogenous intellectual meaning. Basing decisions on a scientific category that is no more stable than sand from the banks of the river Seine can but undermine any final conclusions.

It must also be said that the scepticism towards aid conditionality that now reigns in a certain number of development agencies and intellectual circles is quite dumbfounding for those who have been part of the great development adventure over

the last twenty years. The critique of conditionality, which supposedly "fails" to influence our partners, first of all neglects a point of the critique of conditionality itself. Fortunately, our partners have not implemented some of the conditionalities of the Bretton Woods institutions! I believe I can speak about this with some sense, as I was once a part of these institutions, and we work with them on a daily basis.

As for countries' economic policies, there is so much disagreement, ambivalence and uncertainty (some fundamental reforms have resulted in bitter failure) about what is correct, that one cannot persistently argue that the conditionalities were not implemented in order to explain that certain countries are bad "performers" or, again, that conditionality "did not succeed". The nature of conditionality means that it also needs a long time to set up. It is pathetically naive to think that a fresh young "PhD", who goes straight from a top university to work at the World Bank, the IMF or for the European Union, will, after three or four years, be able to teach the government of Burkina-Faso how to manage a cotton industry that it has been developing successfully for several years, to demand reforms that could possibly lead to the overthrow of the government and bring strife to the countryside, and then say: "That's strange! It's year end and they still haven't implemented these reforms!" At the same time, of course, he is holding to the government's head the gun of possible non-allocation of structural adjustment credits. How can we not understand that the fundamental reforms demanded (even if they are often opportune) need a long time to be implemented? Arguing that reforms are difficult and slow in order to claim that they do not work reveals an absence of lucidity about the nature of reform processes and their political economy.

I would like to mention a third and related point. Sometimes, conditionality is requested by the countries themselves. All those involved concretely in development are familiar with those end-of-day or over-dinner conversations with a minister—little matter from which continent he comes—who says: "Please, maintain your conditionality so I can tell my President that you are really tough. That way, I will be able to complete the necessary fundamental reforms in an honourable fashion." How many reforms have been carried out in difficult conditions, in poor countries, with the World Bank, the IMF or bilateral donors in the role of the "tough guy"! European Union countries may also sometimes be tempted to hold Bruxelles responsible for certain decisions they have to make. Yet, this is political reality. Certainly, we would prefer a world where politicians always tell the truth and are able to see through difficult

reforms but, in fact, we sometimes need “police” and “tough guys” to shoulder some of the world’s failings.

To round off this topic, I would say without hesitation that conditionality has indeed brought about policy changes in Sub-Saharan Africa. Those of us who have been in the cotton or cocoa fields, the factories or the bureaux of finance ministers, who know what the political economy there was like twenty years ago and who see what is being done today, cannot doubt that things have changed dramatically. These changes were made possible thanks to long and patient collaboration with donors, to inter-country exchange, to experience-sharing, negotiations, failures and setbacks… and advances! After my seemingly harsh words about the World Bank and the IMF, I should like to express my greatest respect for the tremendous work carried out by these institutions, without which such transformations—with all their limits, contradictions, deadlocks and slowness—would never have happened. Conditionality was instrumental in these achievements. It was, thankfully, less effective than they would have wished, but fortunately more effective than some economic literature and the institutions themselves today believe. I am therefore struck by the fact that the research work on conditionality does not reflect this reality, thus leading to changes in aid allocation methods that produce inescapable dead-ends and preclude any real restructuring of aid.

To sum up, I would say that changing policy and modifying the rules of the game are the core of the development aid profession. Certainly, when sectors or countries are well-organised (macro-, meso-, and micro-economically), it is positive, useful and advantageous to provide financing that allows an increased volume of public spending or a reduced fiscal effort. It is also essential, however, to transform the realities of public thinking and action. Experience-sharing, political dialogue, education and capacity building constitute a fundamental added value for a great many aid agencies carrying out such programmes. An interim conclusion on this topic would be that conditionality will doubtless be with us for a long time to come, as it would be extremely difficult for us to do without it.

I have now come to the call for support that I mentioned earlier. The challenges lying ahead for development aid policy over the next ten or twenty years are considerable. During the obscure decade of the 1990s, leaders shelved official aid as a political priority (after the fall of the Berlin Wall, the illusion of the Last Man and the End of History

as formulated by Fukuyama). Yet by the end of the decade, it had re-emerged as a major issue and, today, no G7 or G8 summit goes by without ODA being high on the agenda. Budgets materialise slowly and never soon enough. Nonetheless, global policy-making dialogue between aid professionals and world leaders is incomparable with what it was four or five years ago.

In the coming years, three driving forces will affect the development of aid. The first is geopolitical tension. It is interesting to note that most ODA increases over recent years have gone to Iraq and Afghanistan. Rising levels of conflict, crisis management and sometimes prevention issues have drawn heavily on development aid agencies. We see that the portion of official development aid mobilised or "vitrified" by these preoccupations is growing. This is posing problems, a great many of which still have to be thought through.

The relationship between humanitarian and development work, between prevention and management of crises, as well as aid's role in creating global stability, constitute open ground for research. A number of works already exist, but we still have a great deal to learn, and whole fields of knowledge to develop. Without help from economists, sociologists, political scientists and intellectuals in general, agencies like our own will find it very hard to avoid a host of operational traps,

The second driving force for official development aid over the next years is global public goods. Today, all globalisation-related issues have a pivotal North/South component. No major aspect of global regulation or great humanitarian cause—be it finance, trade, environment, health or global warming—remains unaffected by the question of North/South relations. Not a single one of these problems can be resolved without there being, at some point, financial, intellectual or technological transfers between the North and South. These transfers are either necessary to "buy" global agreements, or they are indispensable from the strictly technical point of view of producing a good (the logic of the "weakest link", the "best shot", etc.). Moreover, global public goods is an issue that increasingly affects ODA professionals.

What will we do tomorrow, to complete the Doha Round, if, for instance,

trade liberalisation accelerates to the extent that poor countries lose some of their trade preferences? What will we development agencies do to guarantee that Burkina-

Faso, Nepal or Bangladesh do not become, *in fine*, the victims of international agreements? Every day, we receive questions related to all these matters from our ministers, civil servants, politicians or ONGs. For many of these subjects, the relationships between aid and those who manage global regulation are still hazy. The issues are poorly understood and poorly elaborated in our agencies. Whether at the level of conceptualisation, performance-measuring or operational implications, we still have a long way to go and it will not be possible to move ahead without the scientific community's engagement.

The final driving force of tomorrow's official development aid is the Millennium Development Goals (MDGs). For the first time in its history, the development aid profession has a common frame of reference, a framework for performance on which political leaderships in the North and South have reached a consensus (real or not).

This framework, however, poses considerable problems, as it completely changes the philosophy of aid, insofar as many poor countries find it difficult to imagine that the Millennium Goals are attainable and sustainable without recurrent external financing over extremely long periods. We are therefore completely changing the spirit of official development aid, whose current philosophy reasons as follows: *"We invest from time to time in order to obtain permanent results. Recurring charges from these investments will be dealt with through the tax revenue and fiscal revenue, etc., generated by the positive impact of investments."* This model remains anchored in the current mentality. However, it no longer has anything to do with the model towards which we are currently heading. This tells us: *"In fact, tomorrow, to achieve the Millennium Development Goals, the international community will assume the operating costs of the public services (salaries, administrative costs, etc.), including hospitals, education, road maintenance, etc., in those countries too fragile to meet the costs alone."*

If the international community takes this idea seriously, we are moving toward a massive, long-term undertaking, whose end is difficult to see,—toward a state-controlled partnership between the poor and developed countries. Some would say that this is barely different from the current situation. Yet, in my opinion, there is indeed a difference. Imagine a Sahelian country where official development aid represents 14% of GDP, with tax revenue at around 12-13% of GDP, and the same country with the same rate of fiscal pressure, but with aid amounting to 30-40% of GDP, with teachers, doctors, and

functionaries who know that their salary levels will be decided in Washington, Paris and London, *de facto,* because, globally, the entire running of the state (particularly the social sector) will be covered by international aid. What we are preparing is a revolution. Our leaders have launched this movement, but we are not yet measuring its political, cultural, strategic, economic and financial consequences.

At the core of the MDGs, there are thus a great many implications that have yet to be thought through. Will it take nothing more than a one-off, massive aid effort to generate the private, national and foreign investment that will successfully expand economic growth, so that one day— in twenty, thirty or forty years' time—, recurring costs will be a thing of the past? Or will the "crowding out" effect of aid play its role to the full and push countries into an absolute, unlimited dependence? Moreover, would this absolute, unlimited dependence pose a real problem? The Creuse region—for our international guests, this is a very beautiful *département* in the centre of France—benefits from massive transfers of state funds, chiefly from the Ile-de-France *département* in which Paris is situated. No one finds it abnormal that a French citizen in the Creuse region benefit from this national solidarity till the end of his/her days.

Are we now preparing the same system of redistribution on an international scale? And is this system bad? Not necessarily, if we imagine that, in world markets, a social-democratic concept of globalisation leads quite logically to the setting up of permanent redistribution mechanisms. What must be done, however, is to think clearly about this, recognise it and state it in these terms.

To end on this point, we can imagine only too well that the issue of the MDGs is leading us toward a new world, in which—and this is the last call for support—we will need all the strength and capabilities of our American, Asian, African and Latin American friends in order to successfully re-address all these problems. Certainly, these are the intellectual challenges that we are facing together.

Publications

Les titres de la collection Notes et Documents
sont disponibles sur le site Internet de l'AFD

All volumes of the Notes and Documents
series are available on line at:

www.afd.fr, Publications

N° 1 : Compétitivité et mise à niveau des entreprises (2003)

N° 2 : Multinationales et développement : le rôle des politiques nationales (2003)

N° 3 : Lutte contre l'effet de serre : enjeux et débats (2003)

N° 4 : Comment financer durablement les aires protégées à Madagascar ? (2003)

N° 5 : Agriculture et commerce : quels enjeux pour l'aide au développement ? (2003)

N° 6 : Efficacité et allocation de l'aide : revue des débats (2004)

N° 7 : Qui mérite l'aide ? Égalité des chances *versus* sélectivité (2004)

N° 8 : Le Cambodge : de l'ère des quotas textiles au libre-échange (2004)
Life after Quotas: A Case Study of the Cambodian Textile Industry (2005)

N° 9 : La Turquie : panorama et répartition régionale du secteur productif (2004)
Turkey: Overview of the Economic Productive Sector and Regional Spread of the SMEs

N° 10 : *Poverty, Inequality and Growth, Proceedings of the AFD-EUDN Conference 2003* (2004)

N° 11 : *Foreign Direct Investment in Developing Countries: Leveraging the Role of Multinationals* (2004)

N° 12 : Libre-échange euro-méditerranéen : premier bilan au Maroc et en Tunisie (2005)

N° 13 : Les Mécanismes de financement de la formation professionnelle : une comparaison Europe – Afrique (2005)
Financing Vocational Training: a Europe-Africa Comparison (2005)

N° 14 : Les Mécanismes de la formation professionnelle : Algérie, Maroc, Tunisie, Sénégal (2005)

N° 15 : Les Mécanismes de la formation professionnelle : Allemagne, Espagne, France, Royaume-Uni (2005)

N° 16 : Le Textile-habillement tunisien et le défi de la libéralisation : quel rôle pour l'investissement direct étranger ? (2005)

N° 17 : Poulina, un management tunisien (2005)

N° 18 : Les programmes de mise à niveau des entreprises : Tunisie, Maroc, Sénégal (2005)

N° 19 : Analyser l'impact d'un projet de microfinance : l'exemple d'AdéFI à Madagascar (2005)

N° 20 : Précis de réglementation de la microfinance, tome I (2005)

N° 21 : Précis de réglementation de la microfinance, tome II (2005)

What is AFD ?

The Agence Française de Développement (AFD) is a major operator of French overseas development assistance (ODA), together with the Ministry of Foreign Affairs and the Ministry of Finance (Treasury). Since its creation in 1941, it has been contributing to the development of more than 80 countries as well as to the promotion of French overseas territories. As a financing institution, AFD supports economic, social and environmental projects, with a wide range of instruments from grants to concessional or market conditions loans. Its field of intervention covers productive projects in agriculture, industry or services, either public or private; infrastructures; urban development; education; health; and environment.

5, rue Roland Barthes - 75598 Paris cedex 12
Tél. : 33 (1) 53 44 31 31 - Fax : 33 (1) 44 53 99 39 - www.afd.fr

Création et réalisation Vif Argent Communication - 92200

Imprimé en France par Ferréol (Lyon) - Décembre 2005
Dépôt légal : 4e trimestre 2005