

Working Paper

Improving Dibao Monitoring and Evaluation: Methodologies and Roadmap

Qin Gao

Abstract

Table of Contents

1. Monitoring Management Performance of Social safety net Programs	4
1.2 Definitions and key aspects of management performance.....	4
1.2 Methodologies and evidence around the world	5
1.3 Future directions in monitoring management performance of Dibao.....	11
2. Impact Evaluation of Social Safety Net Programs.....	13
2.1 Definition and key aspects of impact evaluation.	13
2.2 Methodologies and data requirements	14
2.3 Evidence around the world.....	18
2.4 Gaps and future directions in impact evaluation of Dibao.....	19
3. Roadmap and Technical Recommendations: What Can We Do to Improve Dibao Monitoring and Evaluation?.....	20
References.....	23

Improving Dibao Monitoring and Evaluation: Methodologies and Roadmap

Since its inception in Shanghai in 1993, Dibao has been implemented for over 20 years. It was implemented nationwide in the cities in 1999 and in rural areas in 2007. Currently, it is not only China's primary social safety net program, but also the world's largest such program in terms of beneficiary population. In 2014, the total number of Dibao beneficiaries was 70.84 million, including 18.77 million urban beneficiaries and 52.07 rural beneficiaries. Total Dibao expenditure reached 159.20 billion yuan, with 72.17 billion yuan on urban Dibao and 87.03 billion yuan on rural Dibao (Ministry of Civil Affairs [MOCA], 2015). This report reviews the international literature on monitoring management performance and impact evaluation of social safety net programs around the world and offers a roadmap for improving Dibao monitoring and evaluation through better data and methodology preparations and system building.

1. Monitoring Management Performance of Social safety net Programs

1.2 Definitions and key aspects of management performance.

Measuring management performance focuses on the process of program implementation, financing, service delivery, and administration. Management performance of social safety net programs usually refers to the programs' performance in the following aspects:

Budget and finance: This includes how the budget for social safety net is developed, how the funds are managed and whether there are leakages or misuses, whether inter-governmental departments work collaboratively and effectively to develop the budget and allocate the expenditure, and whether the expenditure as a whole and at per beneficiary level meets the social safety net needs of the country, by localities, and by demographic subgroups.

Benefit adequacy and service delivery: Benefit adequacy refers to the level of benefit thresholds and how it relates to financing and targeting discussed above. Service delivery refers to the effectiveness of each key step in the service delivery process, including reaching out, enrollment and approval process, benefit delivery channel, timing and frequency, etc. Combining benefit adequacy and service delivery, we can detect whether the full amount of entitled benefits is delivered smoothly to the eligible individuals and families. A wide gap between the entitled benefit amount and the actual receipt amount, or the benefit gap, would suggest that the benefit delivery has serious errors. A well benefit-targeted program would concentrate total benefits amongst the target group and deliver the full amounts of entitled benefits to them. Another aspect of benefit delivery performance is the time taken for the benefit to reach the beneficiaries.

Population coverage and targeting performance: Population coverage refers to the share of the population receiving social safety net benefits and its variations by localities as well as urban-rural areas as well as whether such shares match the poverty statistics and actual needs of the population. An examination of the demographic and socioeconomic characteristics of the beneficiaries can

help assess whether social safety net programs indeed reach those who deserve the support. Targeting performance refers to the extent to which the social safety net benefits actually reach its intended target population, and on the reverse side, the extent to which such targets are missed.

Vertical and horizontal coordination: This refers to whether various government departments involved and their local branches are efficient and effective in vertical and horizontal coordination during the planning and implementation of the social safety net programs. One very important element of such coordination is whether the information and funding flows between the vertical and horizontal channels are clear and smooth.

Internal and external audits: It is common practice in developed countries to have a well-designed and implemented system of internal and external auditing to ensure the accountability and effectiveness of the management performance of social safety net programs. It is important to understand whether such systems exist for Dibao and whether they serve the auditing purpose sufficiently and help enhance the overall management performance of the program.

It is important to closely monitor and evaluate the management performance of Dibao and other social safety net programs in order to improve the cost-effectiveness of such programs and ensure that they achieve the intended goal of providing an appropriate and sufficient safety net to people in need.

1.2 Methodologies and evidence around the world

Internationally, social safety net programs can be grouped into unconditional and conditional cash transfer programs. Both are means-tested and target the poor. However, conditional cash transfers (CCT) programs usually make benefit receipt conditional upon recipients' actions in human capital investment (in most cases) such as school enrollment for children and regular doctor's visits or participation in welfare-to-work programs, while unconditional cash transfer (UCT) programs do not have such requirements. UCT programs have a long history and are widespread across developed and developing countries, while CCT programs have been implemented relatively recently and are most popular in Latin American and Caribbean countries. This section reviews the international evidence on monitoring management performance of social safety net programs, aiming to offer useful implications for Dibao.

The majority literature on monitoring management performance of social safety net programs has relied on administrative data to examine the financing, population coverage, and local variations of such programs. In particular, central (in centralized systems such as South Korea) or local governments (in decentralized systems such as the US) have been tracking the funding sources (central vs. local government) and budget sizes, welfare caseload, targeting and benefit delivery performance of such programs using monthly or quarterly aggregate data. In recent years, administrative data are increasingly supplemented by large-scale household survey data that can support more detailed and dynamic analysis of management performance of social safety net programs, especially regarding targeting and benefit delivery performance.

First, with regard to *budget and finance*, based on an extensive review of social safety net programs in 136 countries, World Bank (2015) found that safety net programs are affordable at all levels of income. Low- and middle-income countries devote about the same level of resources to social safety nets (1.5 and 1.6 percent of GDP, respectively), while richer countries spend 1.9 percent of GDP on such programs. Lower-income countries devote a higher share of their social safety net budgets to means-tested, targeted programs, while higher-income countries devote a higher share to universal or other categorical programs. Still, many countries lack cost-efficiency in their social safety net programs, as measured by poverty gap reduction and the cost of social safety net benefits as a percentage of household consumption.

The budget committed to social safety net programs is determined not only by a country's financial capacity but also by its commitment to reducing poverty, especially extreme poverty, and its political will and power to do so. Such political will and power are often deeply rooted in a country's social values and principles, and in particular, how poverty is defined and viewed by the society (Barrientos, 2013; Behrendt, 2002; Eardley et al., 1996; Grosh et al., 2008; Umapathi, Wang, and O'Keefe, 2013). Without considering other factors in program design and implementation, a larger budget committed to social safety net programs is usually associated with better poverty alleviation outcomes. World Bank (2015) argued that the efficiency of spending could be improved by strengthening institutional capacity, coordination, program administration, and evaluation, all of which are important in the case of Dibao in China.

In China, Dibao budget is committed by local governments but heavily subsidized by the central government, especially for the less developed and financially challenged localities. The room for the local governments to set their own Dibao budget is increasingly limited given the strengthened role of the central government to provide regulations and guidelines regarding setting up Dibao thresholds and offering sufficient population coverage (Umapathi, Wang, and O'Keefe, 2013). However, there is notable lack of transparency and accountability in the budgeting process of Dibao, especially its central-local negotiations and budgetary responsibilities. While more expert input has been sought in Dibao's impact evaluation in recent years, there remains little—if any—expert input in the budgeting process for Dibao.

It is possible that the budget and financing of Dibao can be made more transparent and efficient. In this regard, Dibao can learn from other aspects of performance management in the governing process. Wong (2012) mentioned several good examples of local experiments in performance management in selected localities such as Guangdong and Shanghai. For example, in Guangdong province, Nanhai city used a team of expert consultants to review several competing budget proposals to help enhance the appropriateness and efficiency of the budget. This process also pushed government officials to prepare better and more realistic budget proposals with a longer-term financial perspective. The provincial capital city, Guangzhou, and Shanghai implemented an ex post evaluation system of spending programs that provided helpful feedback to the budget appropriations concerning specific departments and programs (Ma and Wu, 2011).

Second, with regard to *benefit adequacy and service delivery*, World Bank (2015) discovered that, within a given budgetary framework, many countries try to strike a balance between expanding

population coverage and providing more adequate transfers to a smaller group of the poor. In reality, it is often a trade-off determined by each country's social value system and political environment. Across countries, the take-up rate for social safety net is often low compared with that for other types of social benefits. Benefit generosity is also lower in developing than developed countries, with the median program adding only 10-20% to pre-transfer consumption of their beneficiaries in developing countries (Grosh et al., 2008). The low take-up rate and the low benefit generosity have both limited the anti-poverty effectiveness of the social safety net programs.

Benefit adequacy and service delivery are also closely tied to the means testing procedures adopted by each country and their administrative systems. Some have stricter and more costly means testing procedures than others. For example, among the OECD countries, some (e.g. Australia and UK) have integrated national administrative schemes with common rules of eligibility and payment levels, while others (e.g. Italy, Norway, and Switzerland) have adopted decentralized systems where the local governments are wholly in charge of administration, and funding is often split between central and local governments.

Brazil's Bolsa Família Program uses unverified means testing at the municipal level to determine eligibility (Soares, Ribas, and Osório, 2010). While self-reported income is not verified, this system is supported by two other verification methods. First, the application form also gathers information on family consumption, which is then used to cross check self-reported income. In practice, a double check is warranted when the reported consumption is 20% greater than reported income. Second, Brazil has a federal database that contains information on formal sector workers' employment status and earnings, which is used for cross checking self-reported income. As the database is preexisting, it does not generate additional cost for use or stigma associated with means testing as a way to verify income for determining eligibility for Bolsa Família Program.

In the case of Dibao, strict means testing is involved across localities, but local MOCA branches and their officials have lots of room for deciding how to carry out the means testing procedures. Some localities have requirements on work ability, asset ownership, and family formation, while some others rely on income information more heavily (Solinger, 2010, 2011; Solinger and Hu, 2012). Thus far, there has been no systematic research about the decentralized means testing procedures used and how they affect the benefit adequacy and service delivery in Dibao. Existing qualitative data offer glimpse into the extent of such variations and how it might be linked to the determination of benefits received by Dibao applicants (Han, 2012; Solinger, 2010, 2011; Solinger and Hu, 2012), but more systematic studies, particularly in collaboration with selected local MOCA offices, would help us gain knowledge and insights into this important issue.

Existing studies revealed notable benefit delivery gaps in Dibao (i.e., the difference between entitled Dibao benefit and the actual receipt amount), with the entitled amounts not always fully delivered to eligible beneficiaries. For example, using NBS Urban Household Survey data from 35 large cities in 2003 and 2004, Wang (2006) found that average annual income of eligible families was 893 yuan lower than the Dibao line in 2004, and the average Dibao benefits of eligible families was only 273 yuan, yielding a benefit gap of 620 yuan. Using the China Household Income Project (CHIP) 2002 and 2007 data, Gao and colleagues (2009; 2015) found that the Dibao

benefit gap remained substantial in both years, accounting for about one quarter of the potential full post-Dibao income of all eligible families. Among those eligible for Dibao, the benefit gap among Dibao recipients more than halved from 683 yuan in 2002 to 326 yuan in 2007, indicating better service delivery among the beneficiaries during this period. However, for the eligible non-recipients, the benefit gap was wider in 2007 than in 2002, suggesting that this group needs not only to be better targeted but also more effective benefit delivery in future.

Third, with regard to *population coverage and targeting performance*, despite the multiple efforts to enhance means testing and targeting performance, most poor families around the world still remain outside the social safety net programs, especially in lower-income countries. Across low- and lower-middle-income countries, only about one quarter of the poorest quintile is covered by social safety net programs, while that share is higher at 64% among upper-middle-income countries. This under coverage is especially serious in Sub-Saharan Africa and South Asia, where most of the global poor live, and more challenging among the urban poor in developing countries than in rural areas. CCT programs typically have better targeting performance than UCTs, with nearly 50% of the benefits going to the poorest income quintile (World Bank, 2015).

Targeting performance is closely related to the rules and procedures adopted by each country and program for means testing. Usually, more stringent means testing means better targeting performance, but at higher administrative and social costs. It has higher administrative cost as more personnel and procedures are involved in the means testing process; it has higher social cost as such a stringent process usually involves more disclosure of private information (in terms of both income and other family circumstances such as health condition and family relations), greater peer pressure, and more stigma and shame for the applicants.

Another important cost could be the lower poverty reduction effects due to better targeting and narrower population coverage as a result of more stringent means testing. As Soares, Ribas, and Osório (2010) revealed, cash transfer programs often face a trade-off between better targeting and extending coverage. Contrasting Brazil's Bolsa Família Program with Mexico's Oportunidades Program, both large CCT programs, they found that Oportunidades had better targeting performance than Bolsa Família, but at the price of covering fewer poor households and having smaller overall anti-poverty impact. Indeed, as Ravallion (2009) pointed out, good targeting coupled with small population coverage often leads to limited poverty reduction impact, an important trade-off that is worth more research attention and policy debates.

In the international literature on social safety net programs as well as in the literature on Dibao, the most widely used approach to measure targeting performance has been to use large-scale household survey data to estimate leakage and mis-targeting rates, which respectively reflect the exclusion and inclusion errors in the population coverage of the social safety net programs (see, for example, Du and Park, 2007; Han and Xu, 2014; Gao, Garfinkel, and Zhai, 2009; Gao, Zhai, Yang, and Li, 2015; Gustafsson and Deng, 2011; Soares, Ribas, and Osório, 2010). Specifically, the leakage rate (i.e., exclusion error rate) refers to the proportion of those who are eligible for the benefits but do not receive the benefits. The mis-targeting rate (i.e., inclusion error rate) measures the proportion of ineligible recipients out of all recipients. Ravallion, Chen, and Wang (2006)

argued that, despite the existence of targeting errors in Dibao, based on international standards, such targeting performance is quite good for a means-tested public assistance program.

Another approach to measure targeting, however, is much less used and can add important insights into the targeting performance of various social safety net programs. This approach aims to capture the share of cash transfers going to the poorest income groups (usually deciles or quintiles) as well as the concentration index of the cash transfers (Coady, Grosh, and Hoddinott, 2004; Ravallion, 2009). If targeting is effective, then the share of cash transfers going to the bottom income groups should be higher than the higher income groups, and the concentration index should be more negative, indicating more progressive redistribution of the cash transfers.

In evaluating the targeting performance of Dibao, three existing studies have used this approach, providing a more nuanced understanding of targeting than simply estimating leakage and mis-targeting rates (Du and Park, 2007; Han and Xu, 2014; Gao and Riskin, 2009). For example, Han and Xu (2014) used survey data among 9,107 rural households in five central and western provinces and found that, among the total Dibao benefits delivered to recipient families, only 28% went to those whose family income was below the local Dibao line. The percentage increased to 57% if the national rural poverty line of 2,300 yuan per capita per year was used. Based on this analysis, rural Dibao had serious targeting errors that can be addressed in the future.

It is important to note that, the detailed and accurate analysis of population coverage and target performance of Dibao, as is the case for any other social safety net programs, relies on well-designed, large-scale household survey data collection that is able to capture the income dynamics and social welfare participation of poor families. Longitudinal data collection efforts would further facilitate more in-depth understanding of the welfare entry and exit of these families.

Fourth, with regard to *vertical and horizontal coordination*, some countries have been using key macro indicators to measure the management performance of social safety net programs. In particular, different indicators and measures have been adopted to ensure efficient and effective coordination among various government departments involved and their local branches and smooth information and funding flows between the vertical and horizontal channels.

For example, to monitor and evaluate the Bolsa Família Program, the large CCT program implemented in a decentralized context, Brazil adopted a four-level performance measurement system to assess on-going program execution (short-term indicators), implementation processes, outputs and results (medium-term indicators), and impacts (long-term indicators). Across the four levels, the program managers keep track of a list of key physical and financial program indicators on a monthly basis. These include financial or payment indicators, physical indicators of coverage, targeting and registration indicators, and physical indicators of compliance with education and health conditionalities (given that the Bolsa Família is a CCT program). Two databases—one for the financial indicators and the other for the physical indicators—are set up to store the data cumulatively, with cross-referencing across the two databases possible. The data are collected in collaboration with various municipalities. These data are not only important for tracking the management performance of the Bolsa Família Program from multiple dimensions, but also

enables more accurate and in-depth impact evaluation of the program in the diverse and dynamic local contexts (Lindert et al., 2007).

Brazil's Ministry of Social Development also adopted a Decentralized Management Index (DMI) to monitor and evaluate the quality of program implementation in each municipality. It collects local data on key indicators of program implementation across municipalities, stores the data centrally, and makes the information transparent. More importantly, this system serves as the basis for offering performance-based financial incentives (i.e., providing administrative cost support) to promote quality in municipal implementation (Lindert et al., 2007). This offers a useful model for adaptation in Dibao and social safety net programs in other countries facing similar challenges in decentralized contexts.

It is important to note that the Bolsa Família Program has played a unique unifying role in integrating various Brazilian social policies and programs. Horizontally, it has integrated several federal CCT programs as well as linked with various complementary services and programs. Vertically, it has integrated various subnational CCT programs. These integrations have led to greater coherence of Brazil's social policy agenda and facilitated the building of a more comprehensive and effective social safety net for the country (Lindert et al., 2007).

In the case of Dibao, while there exists rich administrative data on Dibao expenditures and number of beneficiaries, both nationally and by province/city/county, more analysis can be done in collaboration with key government departments involved to help better understand the management performance as well as impacts of Dibao, both at the national and local levels and generate useful policy implications.

Schreyer and Holz (2005) and Wong (2012) highlighted two system challenges that affect data quality and coordination of statistical reporting in China: fragmentation of the reporting system across various government departments and the National Bureau of Statistics (NBS)'s weak control over reporting by subnational offices. Similar challenges face the statistical reporting of Dibao.

Currently, all statistics about Dibao are published by MOCA, including its annual and quarterly expenditures and number of beneficiaries at the national and local levels. Indeed, the MOCA is ahead of many other ministries in releasing Dibao data publicly and in a timely fashion. However, relatively little existing data or research about how the budgeting for Dibao is decided and whether scientific evidence is considered in the budgeting process. To strengthen the administrative data availability and enhance its accountability, crosschecked and jointly released budgeting data on Dibao from both MOF and MOCA would be a good starting point. This system can also help shed light on how the central-local financing responsibilities for Dibao are decided and whether it is done with support of scientific evidence and in response to any particular reality challenges. More involvement of the NBS would also help improve the professionalism of Dibao data reporting and solve possible discrepancies across various channels of reporting.

Wong (2012) noted that China's process of building a performance-oriented management system has been fragmented among various central and local government departments and agencies, with lack of coordination a main problem. However, coordination does not happen automatically. There needs to be a built system to facilitate and monitor coordination. In the case of Dibao, it is important to understand and help improve the vertical and horizontal coordination among key government departments involved in its design, planning, and implementation. This can be achieved by jointly developing frameworks and key indicators that are accepted and adopted by all departments involved.

Fifth, with regard to *internal and external audits*, while many developed countries have a well designed and implemented system of internal and external audit to ensure the accountability and effectiveness of the management performance of social safety net programs, such systems are largely lacking in developing countries. In the case of Dibao, there is an urgent need to establish an audit system to improve the transparency, accountability, and efficiency of its management performance. This includes both internal audits within the various levels and branches of the finance and civil affairs offices as well as external audits from other independent audit offices such as the National Audit Office (NAO) and its local subordinates or non-profit audit agencies.

Wong (2012) showcased how the NAO's work, especially its criticisms of poor budgeting practice and loose financial management, prompted immediate changes in budgeting procedures and accelerated budget management reforms by the Ministry of Finance (MOF). However, while noting the remarkable success in being able to publish annual audit reports and playing a more powerful monitoring role achieved by the NAO during the past 15 years, Wong (2012) also pointed out the very limited progress in audit at the subnational level. Most audit findings at the local level have not been published because of continued political interference in audit selection and disclosure. Another constraint is the limited resources and inadequate staff capacity at the local level. Most local audit offices have low staffing levels and the staff lack the skills required to fully carry out their audit tasks. These constraints lead the subnational audits to largely focus on financial compliance but ignore performance monitoring and evaluation.

A strengthened role of the NAO, especially at the subnational levels, is crucial for pushing for more transparency, accountability, and efficiency in government performance monitoring and evaluation in all domains and requires stronger support from the central government. Specifically, the management performance of Dibao would benefit strongly from an effective audit system, given its interdepartmental nature in budgeting and financing across MOF and MOCA and its decentralized implementation at each subnational level. Local NAOs can be strengthened to play a more active audit role in enhancing the transparency and accountability of Dibao at the local levels. As Dibao directly aims to improve human livelihood and well-being, it is also possible to engage some non-profit audit agencies to carry out local level audit duties to help enhance the management performance of Dibao.

1.3 Future directions in monitoring management performance of Dibao.

Based on the international literature reviewed above, the monitoring of the management performance of Dibao can be improved through the following data and methodology preparations and system building.

First, it would be really useful for the MOCA, MOF, and NBS to *jointly establish a management performance monitoring system on Dibao*. The coordinated efforts among these three key players in Dibao's design, budget and financing, implementation, and data collection would help make the database of great value and good quality. Learning from Brazil's example, a set of key indicators can be established and adopted, and both the national and subnational branches of these ministries/bureau should be involved in data collection, sharing, and crosschecking. MOF can be responsible for the financial indicators, while MOCA can be responsible for the implementation indicators. Of course, these efforts should be concerted and coordinated so that the monitoring system serves the purpose of improving Dibao's management performance instead of adding additional burden to these ministries. The data collection and management expertise of the NBS can help improve the data quality and sharing. This system can help improve the transparency and efficiency of the budget and financing of Dibao and enhance the vertical and horizontal coordination among the various government departments involved and their local branches.

Second and closely related, it is important to *establish an effective audit system* that would help monitor and enhance the accountability of Dibao's budget, financing, and implementation. As mentioned earlier, there have been good examples in both the national and local governments in China to establish and utilize the audit systems to improve government performance, especially regarding budgeting. For Dibao, this can be done at both the national and subnational levels. A few local experiments can be a good starting point.

Third, having more large-scale surveys, especially those focusing on Dibao's target population, would help enable more accurate assessments of its targeting and benefit delivery performances. Such data collection efforts can be done jointly by government departments, especially selected local governments, and poverty and social policy scholars, with the support and involvement of important international organizations such as the World Bank. The micro evidence from these survey data need to be compared against the macro evidence from administrative data to evaluate their consistency and possible discrepancies. Such a comparison can help reveal possible gaps and future directions for better monitoring of Dibao's management performance.

Across countries and especially in developed countries, national household survey data have been used to more accurately estimate the targeting and benefit delivery performance of social safety net programs. Such surveys have the benefit of obtaining detailed household income and asset information without direct intrusion from government officials, making them possible to yield more accurate estimates. However, such surveys often suffer from the inconsistency between the survey accounting period (usually in annual terms) and the actual implementation of social safety net programs (usually by month).

Lastly, cost-benefit analysis has been another useful tool to assess the management performance of social safety net programs, focusing on the monetary aspect of such programs while also

considering non-monetary costs and benefits. Such analysis typically assesses the costs and benefits of social safety net programs to the beneficiaries, the non-beneficiaries, and the whole society.

2. Impact Evaluation of Social Safety Net Programs

2.1 Definition and key aspects of impact evaluation.

Different from measuring management performance, impact evaluation aims to examine the outcomes from program intervention. Specifically, impact evaluation refers to the evaluation of the impacts and consequences of social safety net programs on various outcomes at the aggregate and individual level. Recent decades have seen a global trend in adopting this approach to offer scientific evaluations of the net effectiveness of social safety net programs in terms of their policy design, reform and formulation, and long-term impacts. The literature has focused on the following aspects in the impact evaluation of social safety net programs:

Income and poverty: The direct goal of social safety net programs around the world is usually to lift the income level of poor families through cash transfers and alleviate poverty. Thus income and poverty have been the most examined outcomes in impact evaluation of social safety net programs. The literature has investigated the effectiveness of social safety net programs in improving income and reducing poverty both at the aggregate population level and the individual beneficiary level.

Wealth and assets: Some social safety net programs explicitly or implicitly aim to help low-income families to build wealth and assets that may help these families build a financial cushion for the future and increase their chance of escaping from poverty. In particular, some CCT programs have built-in goals of helping families set aside money as savings or purchase certain goods that could be considered assets. A series of asset-building programs across the developing and developed world such as child savings accounts explicitly aim at helping low-income families build assets (McKernan & Sherraden, 2008; Sherraden, 2005). The literature has examined both the short- and long-term effects of these programs.

Family consumption and investment in health and education: How do recipient families spend the cash transfer money? Do they use it to meet basic survival needs (i.e., paying for food, shelter, clothing, etc.) or invest in human capital (i.e., paying for health and education)? If social safety net indeed enables recipient families to invest in human capital, then it might have positive long-term effects in helping these poor families and their children to escape poverty, have better life opportunities, and become self-sufficient. If, however, the majority of the cash transfers are spent on meeting day-to-day survival needs, then they are merely functioning as a basic safety net and lack long-term positive consequences.

Welfare to work: Many social safety net programs have explicit requirements for beneficiaries to participate in skills and job training programs to help enable them to move from welfare to work and achieve long-term self-support instead of relying on welfare. The evaluation of welfare-to-work effectiveness usually includes the examination of the work capabilities and activities as well

as the extent of welfare dependency among beneficiaries, barriers and facilitators for the recipients' work efforts at the individual, family, community, and policy levels, and the impacts of various welfare-to-work initiatives and programs.

Behavioral and subjective outcomes such as social activities, time use, happiness, and overall life quality: Another important yet understudied aspect of the impacts and consequences of social safety net programs are behavioral and subjective outcomes of the beneficiaries. Most importantly, given the strict means testing nature of most social safety net programs, it is important to understand whether receiving such benefits is associated with some unintended adverse behavioral and subjective consequences such as reduced labor force participation, less active social interactions, increased idle time, and lower happiness level as well as overall life quality. If this is indeed the case, then these important negative outcomes cannot be ignored in the policy design and implementation of social safety net programs even if they are effective in reducing poverty and increasing human capital investment. Conversely, participation in social safety net programs may help promote human development activities such as education and healthcare, especially CCT programs that specifically require such activities.

Taken together, a full evaluation of the possible impacts of the social safety net programs on the various aspects listed above would help provide a comprehensive picture of the effectiveness and consequences of such programs and improve the future design, implementation, and evaluation of the programs.

2.2 Methodologies and data requirements

Impact evaluation of social safety net programs remains a challenging yet rewarding task. It is challenging because social safety net is intertwined with the broader political, social, and economic contexts as well as other social welfare programs on the one hand and all aspects of people's lives on the other. It is rewarding because evaluation evidence is policy relevant and can play important roles in shaping policy debates and directions. Across the developed and developing worlds, impact evaluation of social safety net programs has garnered interests and efforts by researchers, policymakers, NGOs, and government agencies.

Broadly, *four types of research designs have been used in the impact evaluation of social safety net programs, each with its unique data requirements and strengths and limitations.* The first three rely on quantitative methods, using experimental or observational design and statistical analysis, and aim to tease out the possible causal effects of participation in social safety net programs. The last one uses qualitative, ethnographic approaches to understand the life experiences of social safety net participants and offer policy lessons about what might have worked and what might have not.

First, the most rigorous method used in the impact evaluation of social safety net programs is experimental studies. Because CCT programs make receipt of social safety net conditional on certain behaviors such as school enrolment and regular doctor's visits and can easily manipulate the intervention, experimental studies are used much more widely in the evaluation of CCT rather

than UCT programs. Consequentially, the impacts of CCT programs have been systematically examined using rigorous experimental and quasi-experimental designs and pre- vs. post-intervention data collections. The experimental and quasi-experimental designs help effectively address the issue of selection bias and establish the possible causal effect of such programs. Such rigorous designs are often supported by a close collaboration between researchers and program implementers that helps ensure fidelity to the design and monitoring of the short- and long-term effects of these programs (Banerjee and Duflo, 2009).

Typically, CCT programs are designed to provide cash transfers to one group and compare the outcomes of this group to those of a control group. Members of the control group are equally eligible for the transfers but do not receive them. In a true experiment, the two groups are decided by random assignment and thus are similar in nearly all characteristics except for whether they receive the intervention. Whenever possible, researchers often collect pre- and post-intervention data to compare the baseline information and post-intervention outcomes to eliminate the influence of other factors and identify the true effects of the social safety net programs (Banerjee and Duflo, 2009; Rawlings and Rubio, 2005; World Bank 2015).

When random assignment is not possible, researchers often adopt quasi-experimental designs, which identify a comparison group that does not receive the intervention but is very similar to those who do. A frequently used approach for identifying this group is the waitlist approach. In other words, people who apply for the benefit, who are qualified for it, but are placed on a waitlist due to the lack of quota to accommodate them, would serve as the comparison group. The only difference between this and the treatment or intervention group is that they have submitted their applications somewhat slower than the intervention group. One drawback of this approach is that those in the comparison group might be somewhat less motivated to apply, or it is less convenient for them to apply, than those in the intervention group. However, without random assignment to make a true experiment possible, this type of quasi-experiment is typically the next best choice for obtaining rigorous estimates.

In both experimental and quasi-experimental designs, another strength is that it is possible to examine not only the effects of receiving social safety net, but also the possible effects of various aspects of receiving social safety net, such as its timing (e.g., beginning of month or year, harvest season, new semester), dosage (i.e., amount of benefits), target (e.g., mother, father, child, the whole family), and format (e.g., cash, in-kind, combination of the two). This is especially important for testing and understanding which of these aspects can maximize the positive effects and minimize any possible negative effects of the social safety net program.

Experimental and quasi-experimental designs, however, are much less used in evaluating the impacts of UCT programs. As means testing is usually the criterion for deciding UCT program eligibility, it is often infeasible to use random assignment or even the waitlist approach in implementing such programs. In the absence of experimental designs, quantitative researchers rely on two other methodological strategies discussed below to tease out the possible causal effect of social safety net programs. These two strategies can be used separately but are often used in combination to produce more rigorous estimates when possible.

Second, one of the main strategies used by researchers is to conduct longitudinal studies that measure multidimensional outcomes, with a special focus on child outcomes. Longitudinal studies follow a cohort of people overtime to track changes and detect possible causal links. Often, researchers start with baseline measurements and collect data again at multiple subsequent time points. Data are collected on demographic and socioeconomic characteristics as well as multidimensional outcomes. Importantly, data are collected about the timing and amounts of social safety net receipts of individuals and families and then statistical analyses are conducted to detect the possible effects of social safety net programs. The most commonly used statistical analysis is regressions with rich controls, followed by survival analysis that takes advantage of the longitudinal nature of the data.

Longitudinal studies need to be well designed from the outset so that all key variables are included, especially variables concerning the various aspects of social safety net programs as well as participation in other social welfare programs and outcome variables of interest such as income, poverty, human development, and behavioral and subjective outcomes. Such studies also tend to be costly as it takes months or years to complete the data collection and administrative costs related to tracking people overtime are high. However, longitudinal studies have the unique strengths of being able to establish firm temporal order, which is essential for estimating causality, and eliminating other factors that are also tracked overtime as potential causes of certain effects.

The impacts of most UCT programs in rich nations have been systematically and thoroughly examined using rich longitudinal studies. For example, in the US, the possible effects of various welfare programs have been studied using multiple longitudinal data sources such as the Consumer Expenditure Survey, Current Population Survey, Fragile Families and Child Well-being Study, National Longitudinal Survey of Youth, Survey of Income and Program Participation, and Welfare, Children, and Families, a Three-City Study (Boston, Chicago, and San Antonio). These datasets have provided a rich resource that enabled the rigorous evaluation of the U.S. welfare programs on multidimensional outcomes. Longitudinal data collection efforts are new in China. One promising example is the China Family Panel Studies. Such studies need to be designed to include the necessary measures on welfare broadly and social safety net in particular so that their impacts on different outcomes can be evaluated.

Third, in both longitudinal and cross-sectional studies, researchers also adopt multiple advanced statistical methods to address selection bias and achieve more robust estimates of the causal effects of social safety net programs. Without experimental design, one constant challenge in the impact evaluation of social safety net programs is to identify a plausible comparison group of non-beneficiaries who are similar to the beneficiaries in all other regards except for benefit receipt. Several statistical methods such as difference-in-differences, propensity score matching, and regression discontinuity have been used to address selection bias and approximate the logic of experiments to estimate the effects of social safety net programs.

Specifically, ***difference-in-differences*** compares the outcomes of two groups for two time periods, with a policy intervention dividing the two time periods into pre- and post-treatments. The

treatment group is subject to the intervention while the control group is not. Each group's pre- vs. post- difference in the outcome variable is considered the first difference, and the difference of that difference across the two groups is considered the second difference, hence difference-in-differences (Ashenfelter and Card, 1985). This method requires the two groups to be very similar so that the control group serves as the counterfactual of the treatment group. Empirical studies need to first establish the similarity between the two groups before applying this method. The strength of this method is that it removes biases in estimating the possible effects of the policy intervention—in this case the social safety net program—that could be due to fundamental differences between the two groups as well as other contextual factors. The main challenge of this method remains making sure the two groups are truly similar and comparable and collecting data in both pre- and post-intervention time periods (Kaushal, Gao, and Waldfogel, 2007).

Propensity score matching (PSM) uses regressions to identify non-participants who are similar to participants in a rich set of demographic and socioeconomic characteristics and then compares the two groups to estimate the possible effects of participation of the social safety net programs (Dehejia and Wahba, 1999, 2002; Heckman et al. 1997). A growing set of studies has used this method across countries to provide impact evaluation of social safety net programs (e.g., Himaz 2008; Jalan and Ravallion 2003), including in the case of Dibao (Gao et al., 2014; Gao et al., 2015a, 2015b; Golan, Sicular, and Umapathi, 2014). The advantages of this method include that it draws inference from only proper comparisons and focuses on the population of interest. When matching is conducted based on a rich set of characteristics, any remaining differences between the two groups are minimal and often statistically non-significant; a balance check between the two groups can provide precise information about how similar the two groups are once matched. The major drawback of this method is that the matching procedure can only account for observable factors available in the dataset, which are never exhaustive and thus prohibits a solid conclusion on causality.

Regression discontinuity estimates the causal effects of policy interventions by relying on a cutoff or threshold above or below which the intervention is assigned and comparing the outcomes of observations lying closely on either side of the threshold. For means testing social safety net programs, the cutoff point is often the income threshold used by the program to determine benefit eligibility. Since many such programs have mis-targeting errors, it is very likely in practice that some families close to the threshold receive the benefit while some others do not. Thus, the non-recipients serve as a natural comparison group for the recipients given that they are all very close to the threshold. This method relies on the assumption that assignment to the treatment or intervention around the threshold is random, which may or may not be the case in reality and is hard to establish theoretically or empirically. Its strength is that confounding factors can be eliminated and observations closer to the threshold can be given more weight than those farther from the threshold. Its limitation is that it only focuses on a subsample that is close to the threshold and thus results may not be generalizable to the broader population (Imbens and Lemieux, 2008; Thistlethwaite and Campbell, 1960).

Fourth, qualitative and ethnographic studies offer insights into the daily lives and personal experiences of social safety net recipients, which supplement the quantitative impact evaluations

and help assess the actual influences of the programs on people's lives. Often, interviewees and participants of these studies are recipients themselves and program administrators who can offer insights into what have worked and what need to be improved or changed. This method requires lengthy fieldwork and dedicated researchers who can gain the trust of the study participants. Increasingly, researchers try to combine using quantitative and qualitative methods to provide more accurate and in-depth impact evaluations of social safety net programs. In the US, two recent books used a combination of these methods to reveal the plight of those in extreme poverty and their interactions with welfare programs (Desmond, 2016; Edin and Shaefer, 2015). In the case of Dibao, Solinger (2010, 2011; Solinger and Hu, 2012) has done extensive research to understand the life situations of Dibao recipients.

2.3 Evidence around the world

The international literature has shown *positive effects of both UCT and CCT programs* on raising income levels and lowering poverty rates among their target populations. Overall, social safety net programs have helped reduce poverty and, to some extent, income inequality, at the aggregate society level. This body of literature has been notable for its growing number of rigorous impact evaluation studies using experimental, quasi-experimental, and non-experimental but advanced statistical methods to sort out the possible causal influence of participating in social safety net programs.

Specifically, among the *UCT programs*, make work pay programs (such as the EITC and Child Care Tax Credit in the US and the Basic Livelihood Security program in South Korea) have shown greater anti-poverty effectiveness than traditional means-testing programs (such as the AFCD/TANF in the US and South Korea's Livelihood Protection System, the predecessor of the BLS program). However, both types of UCT programs have had unintended adverse behavioral effects such as reductions in work efforts and marriage.

The *CCT programs*, by design, have both poverty reduction and human development as their central goals and have focused on developing countries and regions. International organizations such as the World Bank and UNICEF have shown strong interests and offered support to CCT programs. Existing evidence has shown that CCT programs have been effective in both poverty reduction and improving health and education outcomes, especially among poor children. Many CCT programs specify mothers as recipients of cash transfers, which has shown positive effects on better family resource allocation and more investment in children. There is some evidence that CCT programs help improve the behavioral and subjective outcomes of recipients. Cash transfers have also shown to have major positive spillover effects on the local economy of target communities (Rawlings and Rubio, 2005; World Bank 2015).

Existing evidence on the impact evaluation of Dibao has focused on the following sets of outcomes. Notably, no existing work has examined the influence of Dibao on wealth and assets, possibly because it is not expected to have an impact on these two outcomes due to its residual, minimal nature.

Income and poverty: The existing literature has shown that Dibao has made modest poverty reduction impacts, but its anti-poverty effectiveness can be improved by better targeting performance and more sufficient delivery of benefits. Dibao has had greater impact on narrowing the poverty gap and decreasing poverty severity than reducing the poverty rate.

Family consumption and investment in health and education: Dibao has helped improve the overall consumption level of poor families and in particular enabled these families to invest more in health and education. However, this positive effect is more predominant in urban areas. Its boost for education investment is not evident for rural Dibao recipients.

Welfare to work: Some local governments have tried to move Dibao recipients with work abilities from welfare to work by requiring skills trainings or take-up of certain jobs. However, there have been very few studies on the effectiveness of such welfare-to-work programs and the existing evidence does not show very promising effects.

Behavioral and subjective outcomes such as social activities, time use, happiness, and overall life quality: The limited evidence on how Dibao might affect these outcomes shows that, like many similar means-tested programs around the world, Dibao has some unintended negative behavioral and subjective consequences. For example, Dibao recipients are less likely to engage in leisure, enrichment, or social activities, more likely to spend time doing housework or being idle, and more likely to be unhappy.

2.4 Gaps and future directions in impact evaluation of Dibao.

The existing impact evaluations on Dibao mainly rely on three types of data sources: administrative data, qualitative and fieldwork data, and quantitative, large-scale survey data. While these data sources compliment each other, evidence based on large-scale survey data is often more reliable and generalizable than the other two sources. Impact evaluation of Dibao would greatly benefit from rigorous study designs and multi-wave data collections following the successful examples of various international evaluation projects. Specifically, the following gaps are identified in the current impact evaluation of Dibao. Future efforts can be made to address these gaps and improve the quality of impact evaluation of Dibao.

The existing Dibao impact evaluations lack well-designed, longitudinal data collection projects focusing on the poor and near-poor (such as SIPP and 3-city study in the US). Most existing Dibao studies rely on national household survey data such as CHIP and CFPS, which have national representativeness but only contain a small sample size of Dibao recipients. Other surveys focusing on the poor population are often conducted as a one-time cross-sectional study and out of convenience or availability of funding. Future Dibao research would benefit from a well-designed, well-coordinated team effort that contains a national poor and near-poor sample and is of longitudinal nature. It is also important to include all the important outcome variables covered above in any future data collection efforts.

Little attention is paid to child outcomes and mental health needs among the Dibao population, both of which are very important but currently ignored. Children from poor families often suffer from many disadvantages and have little hope escaping poverty, being trapped by intergenerational transmission of poverty. The CCT programs have directly and effectively tackled this problem, while Dibao has paid relatively little attention to this problem in its policy design and implementation (with the exception of offering some educational assistance for some cases). To truly enable these families to move away from poverty, future research and implementation of Dibao should focus on investigating child outcomes as well as ways to improve human capital investment for children. Another important challenge revealed by the existing qualitative evidence on Dibao is the mental health needs and consequences among Dibao recipients. More research and policy attention is needed on this important issue.

The current research and data collection efforts on Dibao impact evaluation are not well coordinated. Over the years, various research teams and local governments have launched their respective Dibao studies, which have generated the growing literature on Dibao. However, moving forward, the improvement of Dibao implementation and evaluation would greatly benefit from coordinated research and data collection efforts among interdisciplinary research teams as well as joint efforts by scholars and various governments. The World Bank and other international organizations such as ADB and UNICEF can play a very positive facilitating and organizing role in promoting this effort. It will help greatly improve the quality of Dibao evaluation and implementation.

3. Roadmap and Technical Recommendations: What Can We Do to Improve Dibao Monitoring and Evaluation?

First and foremost, based on lessons from other countries as well as from the existing evidence on Dibao, we need to ***design and carry out more rigorous, better coordinated, and longitudinal research studies*** focusing on Dibao's target population (including the near-poor as a comparison group as well as potential Dibao beneficiaries) and covering multidimensional outcomes. The lack of such data limits the ability for testing the causal influence of Dibao and making direct policy proposals. As Barrientos (2013) pointed out, very few developing or transitional countries have longitudinal datasets suitable for the evaluation of the causal effects of welfare programs. As data collection efforts in China continue to grow, it is of interest to researchers and policymakers to collect and utilize longitudinal data to evaluate the performance and impact of Dibao in the future.

While it is important and informative to use experimental and quasi-experimental designs to monitor and evaluate the management performance and impact of Dibao, researchers and policymakers need to ***be aware of possible contamination in experimental evaluations and identify good solutions to address such issues***. For example, Soares, Ribas, and Osório (2010) discovered that, in the case of an experimental evaluation in Ecuador, both treatment and control groups were contaminated: 42% of the control group actually received the treatment, while 22% of the treatment group did not actually receive the treatment. Such contamination is usually the result of the difficulties in the coordination between program implementation and the evaluation design.

Many researchers *use statistical techniques—such as instrumental variables, regression discontinuity, difference-in-difference, and propensity score matching—to approximate the logic of experimental design* and sort out the possible causal influence of participating in such social safety net programs (Galasso, 2006; Gao, Wu, and Zhai, 2014; Gao, Zhai, et al., 2010, 2014; Han, Gao, and Xu, 2015; Islam, 2014; Shady and Araujo, 2008). A set of recent studies in both medical and welfare program research show that, propensity score matching—the method challenged most often among these three techniques because the matching can only be done based on observables—yields results that are very close to those based on experimental designs (Diaz and Handa, 2006; Handa and Maluccio, 2008; Kitsios et al., 2015; Zahoor et al., 2015).

It is also important to use multiple data sources at both the national and subnational level to *replicate and verify evaluation findings of Dibao*. Internationally, replication research has been a growing trend in firmly establishing impacts of social welfare programs and offering concrete policy implications (Brown, Cameron, and Wood, 2014). Replication research is especially important for understanding and establishing the influence of Dibao given its urban-rural difference and decentralized implementation.

Second, it is important to *examine multidimensional outcomes beyond poverty, income and family consumption*. Social safety net programs have profound and long-term impacts on participating families. Potential outcomes include but are not limited to 1) multidimensional poverty, including material deprivation and hardships; 2) human development, with a particular focus on education, health, and life opportunities; 3) likelihood for breaking off the intergenerational transmission of poverty; 4) moving from welfare to work and independency; and 5) other behavioral and subjective outcomes such as social participation, self-esteem, optimism, and overall life quality.

Third, future evaluations of Dibao should *have a particular focus on children*. If there is one lesson that we can learn from the successful CCT programs around the world, it is the focus on children's human development in addition to poverty reduction. The international literature has documented the wide and persistent existence of the intergenerational poverty trap, which severely limits the life opportunities of children from poor families. Many countries have launched labor market and welfare-to-work programs to reduce welfare dependency by strengthening self-support through education, job training, and increased labor force participation. There have been local experiments in Dibao on such welfare-to-work programs and initiatives. It is important to fully gather and evaluate these experiments and initiatives and offer constructive suggestions about how to improve such efforts in future. Both evaluation and implementation of Dibao in future should focus on investment in children's human capital and their long-term developmental well-being. Only through improved human capital assets and life opportunities can the children from poor families gain a less dismal future and have their fair share in the society.

Fourth, *joint efforts should be made among interdisciplinary scholars, government officials, and international organizations* to improve the monitoring and evaluation of Dibao. Dibao monitoring and evaluation can greatly benefit from coordinated joint efforts among all who care

about Dibao and wish to improve it. International organizations such as the World Bank, ADB, and UNICEF have all worked toward this direction in collaboration with the Chinese government as well as various scholars. A more coordinated term effort will help offer a breakthrough in Dibao research and policy design and place the Chinese social safety net case more prominently in the global dialogue on welfare, work, and poverty.

Lastly, particular attention needs to be paid to Dibao's localized implementation. Dibao is a national policy, but its assistance standards and implementation are very decentralized with many local variations. It is important to carry out close monitoring and rigorous evaluations on selected, representative local Dibao programs to offer insights into what works, what does not work, and what can work better in future. For this, TANF in the US can serve as a good example. It is similarly very decentralized at the state level and many states—in collaboration with research teams—carried out demonstration and evaluation projects that offered ample research evidence that helped improve its effectiveness not only locally but also nationally. Similar work can be done for Dibao, involving local governments and research teams but with national and international collaboration, oversight, and collective expertise.

References

English Literature:

- Aizer, A., Eli, S., Ferrie, J. & Lleras-Muney, A. (2014). The long-term impact of cash transfers to poor families. NBER Working Paper 20103, <http://www.nber.org/papers/w20103>.
- Ashenfelter, Orley and David Card (1985). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. *Review of Economics and Statistics*, 67(4), 648-660.
- Attanasio, O. & Mesnard, A. (2006). The impact of a conditional cash transfer program on consumption in Colombia. *Fiscal Studies*, 27(4): 421-442.
- Banerjee, Abhijit and Duflo, Esther (2009). The Experimental Approach to Development Economics. *Annual Review of Economics*, 1, 151-178.
- Barrientos, A. (2013). *Social safety net in Developing Countries*. Cambridge: Cambridge University Press.
- Behrendt C (2002). *At the margins of the welfare state: social safety net and the alleviation of poverty in Germany, Sweden and the United Kingdom*. Aldershot, UK and Burlington, VT, Ashgate.
- Blank, R. M. (2009). "What We Know, What We Don't Know, and What We Need to Know about Welfare Reform." In *Welfare Reform and Its Long-Term Consequences for America's Poor*, ed. James P. Ziliak. Cambridge: Cambridge University Press.
- Brown, Annette N., Drew B. Cameron & Benjamin D. K. Wood (2014). Quality evidence for policymaking: I'll believe it when I see the replication, *Journal of Development Effectiveness*, 6(3): 215-235.
- Case, Anne and Angus Deaton (1998). Large cash transfers to the elderly in South Africa. *The Economic Journal*, 108: 1330-1361.
- Coady, D., Margareth Grosh, and John Hoddinott (2004). Targeting of Transfers in Developing Countries: Review of Lessons and Experience. Washington, D.C: World Bank and International Food Policy Research Institute.
- Choi J, Choi J (2007). The effectiveness of poverty reduction and the target efficiency of social security transfers in South Korea, 1999-2003. *International Journal of Social Welfare* 16: 183-189.
- Diaz, J. J., and Sudhanshu Handa (2005). "An Assessment of Propensity Score Matching as Nonexperimental Impact Estimator: Evidence from Mexico's PROGRESA Program." *Journal of Human Resources* 41 (2): 319-345.
- Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448), 1053-1062.
- Dehejia, R. H., & Wahba, S. (2002). Propensity Score-matching Methods for Nonexperimental Causal Studies. *The Review of Economics & Statistics*, 84(1), 151-161.
- Desmond, M. (2016). *Evicted: Poverty and Profit in the American City*. Deckle Edge.
- Eardley T, Bradshaw J, Ditch J, Gough I, Whiteford P (1996). Social safety net in OECD countries, Volume I: synthesis report. UK Department of Social Security Research Report No. 46. London, HMSO. Available at <http://research.dwp.gov.uk/asd/asd5/rrep046.pdf>.
- Edin, K. J., & Shaefer, H. L. (2015). *\$2.00 a Day: Living on almost nothing in America*. Boston & New York: Houghton Mifflin Harcourt.
- Fiszbein, A., Schady, N., Ferreira, F. H.G., Grosh, M., Kelleher, N., Olinto, P. & Skoufias, E. (2009). *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington, D.C: World Bank. <https://openknowledge.worldbank.org/handle/10986/2597>.
- Gao, Q. (2013). Public assistance and poverty reduction: The case of Shanghai. *Global Social Policy*, 13(2): 193-215.
- Gao Q, Garfinkel I, Zhai F (2009). Anti-poverty effectiveness of the Minimum Living Standard Assistance Policy in urban China. *Review of Income and Wealth*, 55(s1): 630-655.

- Gao, Q., Kaushal, N., & Waldfogel, J. (2009). How have expansions in the Earned Income Tax Credit affected family expenditures? In Ziliak, J. (ed.), *Ten Years after: Evaluating the Long-Term Effects of Welfare Reform on Children, Families, Welfare, and Work*. Cambridge University Press.
- Gao, Q., & Riskin, C. (2013). Generosity and Participation: Variations in Urban China's Minimum Livelihood Guarantee Policy. In Kennedy, D. & Stiglitz, J. E. (eds.), *Law & Economics with Chinese Characteristics: Institutions for Promoting Development in the 21st Century*. Oxford University Press.
- Gao, Q., Wu, S. & Zhai, F. (2015). Welfare participation and time use in China. *Social Indicators Research*, 124, 863-887.
- Gao, Q., Yang, S. & Li, S. (2015). Welfare, targeting, and anti-poverty effectiveness: The case of urban China. *Quarterly Review of Economics and Finance*, 56, 30-42.
- Gao, Q., Yoo, J. Y., Yang, S., & Zhai, F. (2011). Welfare Residualism: A Comparative Study of the Basic Livelihood Security Systems in China & South Korea, *International Journal of Social Welfare*, 20, 113-124.
- Gao, Q. & Zhai, F. (2012). Anti-poverty family policies in china: A critical evaluation. *Asian Social Work and Policy Review*, 6(1): 122-135.
- Gao, Q., Zhai, F. & Garfinkel, I. (2010). How does public assistance affect family expenditures? The case of urban China. *World Development*, 38(7), 989-1000.
- Gao, Q., Zhai, F. Yang, S. & Li, S. (2014). Does welfare enable family expenditures on human capital? Evidence from China. *World Development*, 64: 219-231.
- Galasso, E. 2006 "With Their Effort and One Opportunity": Alleviating Extreme Poverty in Chile." Mimeo, Development Research Group, World Bank, Washington, D.C.
- Golan, J., Sicular, T., & Umapathi, N. (2014). *Any guarantees? China's rural minimum living standard guarantee program* (World bank social protection & labor discussion paper 1423). Washington, DC: World Bank.
- Grosh, M. E., C. del Ninno, Tesliuc, E., & Ouerghi, A. (2008). *For Protection & Promotion: The Design & Implementation of Effective Safety Nets*, The World Bank, Washington, D.C.
- Guan X (2005). *Poverty in urban China: an introduction*. Beijing, Chinese Academy of Social Sciences Social Policy Research Center.
- Gustafsson, B. and Deng, Q. (2011), Di Bao receipt and its importance for combating poverty in urban China. *Poverty & Public Policy*, 3(1), Article 10. doi:10.2202/1944-2858.1127
- Han, H., Gao, Q. & Xu, Y. (2015). Welfare Participation and Family Consumption Choices in Rural China. Working Paper.
- Handa, S. and Ichni Maluccio 2008 "Matching the Gold Standard: Comparing Experimental and Non-Experimental Evaluation Techniques for a Geographically Targeted Program." Middlebury College Economics Discussion Paper No. 08-13.
- Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64(4), 605–654.
- Hussain A (2007). Social security in transition. In: Shue V, Wong C, eds. *Paying for progress in China: Public finance, human welfare and changing patterns of inequality*, pp. 96 Hong Kong and New York, Routledge.
- Hoddinott, J. & Skoufias, E. (2004). The impact of PROGRESA on food consumption. *Economic Development and Cultural Change*, 53(1): 37-61.
- Imbens, G. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics* 142 (2): 615–635.
- Islam, T.M. Tonmoy (2014). An exercise to evaluate an anti-poverty program with multiple outcomes using program evaluation. *Economic Letters*, 122: 365-369.
- Jalan, J., & Ravallion, M. (2003). Estimating the benefit incidence of an antipoverty program by propensity-score matching. *Journal of Business & Economic Statistics*, 21(1), 19-30.

- Júnior, Alvaro Luiz Neuenfeldt, Julio Cezar Mairesse Siluk, Marlon Soliman, Elpídio Oscar Benitez Nara, Liane Mahlmann Kipper (2015). Hierarchy the sectorial performance indicators for Brazilian franchises. *Business Process Management Journal*, 21(1): 190–204.
- Kaushal, N., Gao, Q., & Waldfogel, J. (2007). Welfare reform and family expenditures: How are single mothers adapting to the new welfare and work regime? *Social Service Review*, 81(3), 369-398.
- Kitsios GD, Dahabreh IJ, Callahan S, Paulus JK, Campagna AC, & Dargin JM. (2015). Can We Trust Observational Studies Using Propensity Scores in the Critical Care Literature? A Systematic Comparison With Randomized Clinical Trials. *Critical Care Medicine*. 43(9), 1870-9.
- Leung JC (2003). Social security reforms in China: issues and prospects. *International Journal of Social Welfare* 12: 73–85.
- Leung JC (2006). The emergence of social safety net in China. *International Journal of Social Welfare* 15: 188–198.
- Lindert, Kathy, Anja Linder, Jason Hobbs and Bénédicte de la Brière (2007). The Nuts and Bolts of Brazil's Bolsa Família Program: Implementing Conditional Cash Transfers in a Decentralized Context. World Bank Social Protection Discussion Paper No. 709. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/2007/05/7645487/nuts-bolts-brazils-bolsa-familia-program-implementing-conditional-cash-transfers-decentralized-context>.
- Ma, Jun, and Shaolong Wu. 2011. "Performance Evaluation of Fiscal Expenditures: Motivation, Process, Impact and Challenges—Case Study of Guangzhou Municipal Performance Evaluation." In 30 Years of Performance Measurement in Chinese Government, ed. Hanxuan Chen, Jun Ma, and Bao Guoxian. Beijing: Central Compilation and Translation Press.
- McKernan, S.M., & Sherraden, M. (Eds.). (2008). *Asset building and low-income families*. Washington, DC: Urban Institute Press.
- Milligan, K. & Stabile, M. (2011). Do child tax benefits affect the well-being of children? Evidence from Canadian child benefit expansions. *American Economic Journal: Economic Policy*, 3(3): 175-205.
- Ministry of Civil Affairs [MOCA] (2015). Social Service Development Statistical Annual Report 2014. <http://www.MOCA.gov.cn/article/zwgk/mzyw/201506/20150600832371.shtml>.
- Ravallion, Martin (2009). How relevant is targeting to the success of an antipoverty program? *World Bank Research Observer*, 24(2): 205-231.
- Ravallion, M., Chen, S., & Wang, Y. (2006). Does the Di Bao Program Guarantee a Minimum Income in China's Cities? In Lou, W. & S. Wang (Eds.), *Public Finance in China*, World Bank, Washington, D.C..
- Rawlings, L. B. & Rubio, G. M. (2005). Evaluating the impact of conditional cash transfer programs. *The World Bank Research Observer*, 20(1): 29-55.
- Saunders P, Shang X (2001). Social security reform in China's transition to a market economy. *Social Policy & Administration* 35(3): 274–289.
- Schreyer, Paul, and Carsten Holz. 2005. "Institutional Arrangements for the Production of Statistics." In *China in the Global Economy: Governance in China*. Paris: OECD.
- Schady, N. R., and Maria Caridad Araujo (2008) "Cash Transfers, Conditions, and School Enrollment in Ecuador." *Economia* 8 (2): 131-154.
- Sherraden, M. (Ed.). (2005). *Inclusion in the American Dream: Assets, poverty, and public policy*. New York, NY: Oxford University Press.
- Soares, Fábio Veras, Rafael Perez Ribas, and Rafael Guerreiro Osório (2010). Evaluating the Impact of Brazil's Bolsa Família: Cash Transfer Programs in Comparative Perspective. *Latin American Research Review*, 45(2): 173-190.
- Solinger, D. (2010). The urban Dibao: Guarantee for minimum livelihood or for minimal turmoil? In Wu F. (ed.) *Marginalization in China: comparative perspectives*. Palgrave Macmillan.

- Solinger, D. (2011). *Dibaohu* in Distress: The Meager Minimum Livelihood Guarantee System in Wuhan. In Jane Duckett and Beatriz Carillo (eds.), *China's Changing Welfare Mix: Local Perspectives*. London: Routledge, 36-63.
- Solinger, D. and Hu, Y. (2012). Welfare, wealth and poverty in urban China: The *Dibao* and its differential disbursement. *China Quarterly*, 211: 741-764.
- Thistlethwaite, D. and Campbell, D. (1960). Regression-Discontinuity Analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6): 309–317.
- Umapathi, Nithin, Dewen Wang, and Philip O'Keefe (2013). Eligibility Thresholds for Minimum Living Guarantee Programs: International Practices and Implications for China. World Bank Social Protection & Labor Discussion Paper Series No. 1307. <https://openknowledge.worldbank.org/handle/10986/17006>.
- Wang M (2007). Emerging urban poverty and effects of the BLS program on alleviating poverty in China. *China & World Economy* 15(2): 74–88.
- Wong, C. (2012). Performance, Monitoring, and Evaluation in China. World Bank Special Series on the Nuts & Bolts of M&E Systems, Number 23. <https://openknowledge.worldbank.org/handle/10986/17083>.
- World Bank (2015). *The state of social safety nets 2015*. Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/2015/07/24741765/state-social-safety-nets-2015>.
- Zahoor H, Luketich JD, Levy RM, Awais O, Winger DG, Gibson MK, & Nason KS. (2015). A propensity-matched analysis comparing survival after primary minimally invasive esophagectomy followed by adjuvant therapy to neoadjuvant therapy for esophagogastric adenocarcinoma. *The Journal of Thoracic and Cardiovascular Surgery*. 149(2), 538-47.

Chinese Literature:

- Du, Y., & Park, A. (2007). Social safety net Programs and Their Effects on Poverty Reduction in Urban China, *Economic Research*, 12, 24-33. (in Chinese)
- Han, H., & Xu, Y. (2014). The anti-poverty effectiveness of the minimum living standard assistance policy in rural China: Evidence from five central and western provinces. *Economic Review*, 6, 63-77 (in Chinese).
- Han, K. (eds.) (2012). Interviews with Minimum Livelihood Guarantee Recipients in Urban China. Jinan: Shandong Renmin Press. (in Chinese)
- Hu, Xuchang, Gao, Lingzhi, & Cui, Hengzhan (2013). An empirical analysis of the living conditions of urban Dibao families. *Journal of Jinan University Social Science Edition*, 23(2): 58-63. (in Chinese)
- Gao, G., Chen, D. & Cui, H. (2013). Asset cumulation and Dibao assistance: A survey and comparative study of assets among urban Dibao families. *Journal of Nantong University Social Science Edition*, 29(2): 51-63. (in Chinese)